

Technical Report 1325

Tier One Performance Screen Initial Operational Test and Evaluation: 2011 Annual Report

Deirdre J. Knapp, Editor

Human Resources Research Organization

Kate LaPort, Editor

U.S. Army Research Institute

January 2013



**United States Army Research Institute
for the Behavioral and Social Sciences**

Approved for public release; distribution is unlimited

**U.S. Army Research Institute
for the Behavioral and Social Sciences**

**Department of the Army
Deputy Chief of Staff, G1**

Authorized and approved for distribution:



**MICHELLE SAMS, Ph.D.
Director**

Research accomplished under contract
for the Department of the Army by

Human Resources Research Organization

Technical review by

J. Douglas Dressel, U.S. Army Research Institute
Kimberly Owens, U.S. Army Research Institute

NOTICES

DISTRIBUTION: Primary distribution of this Technical Report has been made by ARI. Please address correspondence concerning distribution of reports to: U.S. Army Research Institute for the Behavioral and Social Sciences, ATTN: DAPE-ARI-ZXM, 6000 6th Street (Bldg. 1464 / Mail Stop 5610), Ft. Belvoir, VA 22060-5610

FINAL DISPOSITION: Destroy this Technical Report when it is no longer needed. Do not return it to the U.S. Army Research Institute for the Behavioral and Social Sciences.

NOTE: The findings in this Technical Report are not to be construed as an official Department of the Army position, unless so designated by other authorized documents.

REPORT DOCUMENTATION PAGE					
1. REPORT DATE (dd-mm-yy) January 2013		2. REPORT TYPE Interim		3. DATES COVERED (from. . . to) August 2009 to April 2012	
4. TITLE AND SUBTITLE Tier One Performance Screen Initial Operational Test and Evaluation: 2011 Annual Report				5a. CONTRACT OR GRANT NUMBER W91WAS-09-D-0013	
				5b. PROGRAM ELEMENT NUMBER 622785	
6. EDITORS Deirdre J Knapp (Human Resources Research Organization) ; Kate LaPort (U.S. Army Research Institute)				5c. PROJECT NUMBER A790	
				5d. TASK NUMBER 329	
				5e. WORK UNIT NUMBER	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Human Resources Research Organization 66 Canal Center Plaza, Suite 700 Alexandria, Virginia 22314				8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) U.S. Army Research Institute for the Behavioral and Social Sciences 6000 6 th Street (Bldg 1464 / Mail Stop 5610) Fort Belvoir, VA 22060-5610				10. MONITOR ACRONYM ARI	
				11. MONITOR REPORT NUMBER Technical Report 1325	
12. DISTRIBUTION/AVAILABILITY STATEMENT Approved for public release; distribution is unlimited.					
13. SUPPLEMENTARY NOTES Contracting Officer's Representative and Subject Matter Expert POC: Dr. Kate LaPort					
14. ABSTRACT (<i>Maximum 200 words</i>): Along with educational, medical, and moral screens, the U.S. Army uses a composite score from the Armed Services Vocational Aptitude Battery (ASVAB), the Armed Forces Qualification Test (AFQT), to select new Soldiers. Although the AFQT is useful for selecting new Soldiers, other personal attributes are important to Soldier performance and retention. Based on previous U.S. Army Research Institute (ARI) investigations, the Army selected one promising measure, the Tailored Adaptive Personality Assessment System (TAPAS), for an initial operational test and evaluation (IOT&E), beginning administration to applicants in 2009. Criterion data are being compiled at 6-month intervals from administrative records, from schools for selected military occupational specialties (MOS), and from Soldiers in units. This is the fourth in a series of planned evaluations of the TAPAS, and the first to include in-unit criterion data. Similar to prior research, the cumulative results thus far suggest that several TAPAS scales significantly predict a number of criteria of interest, indicating that the measure holds promise for both selection and classification purposes.					
15. SUBJECT TERMS Personnel, Manpower, Selection and classification					
SECURITY CLASSIFICATION OF			19. LIMITATION OF ABSTRACT Unlimited	20. NUMBER OF PAGES 93	21. RESPONSIBLE PERSON Dorothy Young 703-545-2316
16. REPORT Unclassified	17. ABSTRACT Unclassified	18. THIS PAGE Unclassified			

Standard Form 298

Technical Report 1325

Tier One Performance Screen Initial Operational Test and Evaluation: 2011 Annual Report

Editors

Deirdre J. Knapp

Human Resources Research Organization

Kate A. LaPort

U.S. Army Research Institute

Personnel Assessment Research Unit

Tonia S. Heffner, Chief

U.S. Army Research Institute for the Behavioral and Social Sciences

6000 6th Street, Bldg. 1464

Fort Belvoir, Virginia 22060

January 2013

**Army Project Number
622785A790**

**Personnel Performance
and Training Technology**

Approved for public release; distribution is unlimited.

ACKNOWLEDGEMENTS

There are individuals not listed as authors who made significant contributions to the research described in this report. First and foremost are the Army cadre who support criterion data collection efforts at the schoolhouses. These noncommissioned officers (NCOs) ensure that trainees are scheduled to take the research measures and provide ratings of their Soldiers' performance in training. Those Army personnel who support the in-unit data collections are also instrumental to this research program. Thanks also go to Mr. Irwin Jose and Ms. Sharon Meyers (ARI) and Mr. Doug Brown, Ms. Charlotte Campbell, and Dr. Karen Moriarty (HumRRO) for their important contributions to this research effort.

We also want to extend our appreciation to the Army Test Program Advisory Team (ATPAT), a group of senior NCOs who periodically meet with ARI researchers to help guide this work in a manner that ensures its relevance to the Army and help enable the Army support required to implement the research. Members of the ATPAT are:

MAJ PAUL WALTON
CSM MICHAEL COSPER
CSM BRIAN A. HAMM
CSM JAMES SHULTZ
SGM KENAN HARRINGTON
SGM THOMAS KLINGEL
SGM(R) CLIFFORD MCMILLAN
SGM GREGORY A. RICHARDSON
MSG THOMAS MORGAN
SSG THOMAS HILL
SFC APRIL HANSBERRY
SFC WILLIAM HAYES
SFC KENNETH WILLIAMS
MR. JAMES LEWIS
MR. WILLIAM PALYA
MR. ROBERT STEEN

TIER ONE PERFORMANCE SCREEN INITIAL OPERATIONAL TEST AND EVALUATION: 2011 ANNUAL REPORT

EXECUTIVE SUMMARY

Research Requirement:

In addition to educational, physical, and moral screens, the U.S. Army relies on a composite score from the Armed Services Vocational Aptitude Battery (ASVAB), the Armed Forces Qualification Test (AFQT), to select new Soldiers into the Army. Although the AFQT has proven to be, and will continue to serve as a useful metric for selecting new Soldiers, other personal attributes, in particular non-cognitive attributes (e.g., temperament, interests, and values), are important to entry-level Soldier performance and retention (e.g., Campbell & Knapp, 2001; Ingerick, Diaz, & Putka, 2009; Knapp & Heffner, 2009, 2010; Knapp & Tremble, 2007). Based on previous ARI research, the Army selected one particularly promising measure, the Tailored Adaptive Personality Assessment System (TAPAS), as the basis for an initial operational test and evaluation (IOT&E) of the *Tier One Performance Screen* (TOPS). The TAPAS capitalizes on the latest in testing technology to assess motivation through the measurement of personality characteristics.

Procedure:

In May 2009, the Military Entrance Processing Command (MEPCOM) began administering the TAPAS on the computer adaptive platform for the ASVAB (CAT-ASVAB) at Military Entrance Processing Stations (MEPS). To evaluate the TAPAS, outcome (criterion) data are being collected from Soldiers at multiple points following accession.. Specifically, initial military training (IMT) criterion data are being collected at schools for Soldiers in eight military occupational specialties (MOS). Project teams also are collecting criterion data from Soldiers (regardless of MOS) in their units in multiple waves of site visits during the course of the IOT&E. The criterion measures include job knowledge tests, an attitudinal assessment (the Army Life Questionnaire), and performance rating scales completed by the Soldiers' cadre members (in IMT) or supervisors (in units). Course grades, completion rates, and attrition status are obtained from administrative records for all Soldiers.

A data file containing TAPAS data collected through September 2011 and criterion data collected through December 2011 is the basis for the analyses documented in this report. It consists of a total of 216,565 applicants who took the TAPAS; 176,467 of these individuals were in the TOPS "Applicant Sample," which was limited to applicants who took pre-August 2011 versions of the TAPAS. The Applicant Sample (used for analysis purposes) also excluded Education Tier 3, AFQT Category V, and prior service applicants. The validation sample sizes are considerably smaller, with the IMT Validation Sample comprising 9,502 Soldiers, the In-Unit Validation Sample comprising 387 Soldiers, and the Administrative Validation Sample (which includes Soldiers for whom we have at least one administrative criterion data element, as well as the Soldiers in the IMT and in-unit validation samples) comprising 69,495 Soldiers.

Our approach to analyzing the TAPAS' incremental predictive validity was consistent with previous evaluations of this measure and similar experimental non-cognitive predictors (Ingerick et al., 2009; Knapp & Heffner, 2009, 2010, 2011). In brief, this approach involved testing a series of hierarchical regression models, regressing each criterion measure onto Soldiers' AFQT scores in the first step, followed by their TOPS composite or TAPAS scale scores in the second step. The resulting increment in the multiple correlation (ΔR) when the TAPAS scale scores were added to the baseline regression models served as our index of incremental validity. Scale-level correlations between the TAPAS scale scores and selected criteria were also examined.

Findings:

Results suggest that the TAPAS holds promise for predicting key criteria of interest. Incremental validity beyond the AFQT is promising, primarily for predicting will-do criterion measures (i.e., those measuring non-technical aspects of Soldier performance, such as effort, peer leadership, and personal discipline). Many of the scale-level coefficients are consistent with a theoretical understanding of the TAPAS scales, suggesting that the scales are measuring the characteristics that they are intended to measure.

Utilization and Dissemination of Findings:

The research findings will be used by the U.S. Army Recruiting Command, Army G-1, and Training and Doctrine Command to evaluate the effectiveness of tools used for Army applicant selection and assignment. With each successive set of findings, the TOPS can be revised and refined to meet Army needs and requirements.

TIER ONE PERFORMANCE SCREEN INITIAL OPERATIONAL TEST AND EVALUATION: 2011 ANNUAL REPORT

CONTENTS

	Page
CHAPTER 1: INTRODUCTION	1
Deirdre J. Knapp (HumRRO), Tonia S. Heffner, Leonard A. White, and Kate LaPort (ARI)	
Background	1
The Tier One Performance Screen (TOPS).....	2
Evaluating TOPS	3
Overview of Report	4
CHAPTER 2: DATA FILE DEVELOPMENT	5
D. Matthew Trippe, Laura Ford, Bethany Bynum, and Karen Moriarty (HumRRO)	
Overview of Process.....	5
Description of Data File and Sample Construction	6
Summary	9
CHAPTER 3: DESCRIPTION OF THE TOPS IOT&E PREDICTOR MEASURES	10
Stephen Stark, O. Sasha Chernyshenko, Fritz Drasgow (Drasgow Consulting Group), and Matthew T. Allen (HumRRO)	
Tailored Adaptive Personality Assessment System (TAPAS).....	10
Description	10
Multiple Versions of TAPAS.....	11
TAPAS Composites	13
Armed Services Vocational Aptitude Battery (ASVAB) Content, Structure, and Scoring ..	13
Summary	14
CHAPTER 4: DESCRIPTION AND PSYCHOMETRIC PROPERTIES OF CRITERION MEASURES	15
Karen O. Moriarty, Thomas Kiger, and Chad Peddie (HumRRO)	
Criterion Measure Descriptions.....	16
Job Knowledge Tests (JKTs)	16
Performance Rating Scales (PRS)	16
Army Life Questionnaire (ALQ)	19
Administrative Criteria	19
Basic Descriptive Statistics and Psychometric Properties for IMT and In-Unit Criteria.....	19
Job Knowledge Tests (JKTs)	21
Performance Rating Scales (PRS)	21
Army Life Questionnaire (ALQ)	24
Administrative Criteria	25
Summary	29

CONTENTS (CONTINUED)

	Page
CHAPTER 5: EVIDENCE FOR THE PREDICTIVE VALIDITY OF THE TAPAS.....	30
Joseph P. Caramagno and Michael J. Ingerick (HumRRO)	
Analysis Approach	30
Findings.....	32
Summary	45
CHAPTER 6: SUMMARY AND A LOOK AHEAD	49
Deirdre J. Knapp (HumRRO), Tonia S. Heffner, Leonard A. White, and Kate LaPort (ARI)	
Summary of the TOPS IOT&E Method.....	49
Summary of Evaluation Results to Date	50
Looking Ahead	50
Predictor Measures.....	50
Criterion Measures.....	51
Analyses	51
REFERENCES	53
APPENDIX A - PREDICTOR MEASURE PSYCHOMETRIC PROPERTIES IN THE APPLICANT SAMPLE.....	A-1
APPENDIX B - CRITERION PSYCHOMETRIC PROPERTIES IN THE FULL IMT SAMPLE.....	B-1
APPENDIX C - CRITERION PSYCHOMETRIC PROPERTIES IN THE FULL IN- UNIT SAMPLE	C-1

LIST OF TABLES

Table 2.1. Full TAPAS Data File Characteristics	7
Table 2.2. Background and Demographic Characteristics of the TOPS Samples	8
Table 3.1. TAPAS Dimensions Assessed	12
Table 4.1. Summary of IMT and In-Unit Criterion Measures	15
Table 4.2. ALQ Likert-Type Scales	20
Table 4.3. Descriptive Statistics and Reliability Estimates for IMT and In-Unit WTBD and MOS-Specific Job Knowledge Tests (JKTs) by Education Tier in the Validation Sample	22
Table 4.4. Descriptive Statistics and Reliability Estimates for IMT Army-Wide and MOS-Specific Performance Rating Scales (PRS) by Education Tier in the Validation Sample	23
Table 4.5. Descriptive Statistics and Reliability Estimates for In-Unit Army-Wide Performance Rating Scales (PRS) in the Validation Sample	25
Table 4.6. Descriptive Statistics and Reliability Estimates for the IMT Army Life Questionnaire (ALQ) by Education Tier in the Validation Sample	26
Table 4.7. Descriptive Statistics and Reliability Estimates for the In-Unit Army Life Questionnaire (ALQ) by Education Tier in the Validation Sample	27
Table 4.8. Base Rates or Descriptive Statistics for Administrative Attrition and IMT Criteria by Education Tier in the Validation Sample	28
Table 5.1. Summary of the Regression Models	31
Table 5.2. Incremental Validity Estimates for the TAPAS and TOPS Composite Scales over the AFQT for Predicting IMT Technical Performance Criteria by Education Tier	33
Table 5.3. Incremental Validity Estimates for the TAPAS and TOPS Composite Scales over the AFQT for Predicting IMT Effort and Disciplinary Incidence Criteria by Education Tier	34
Table 5.4. Incremental Validity Estimates for the TAPAS and TOPS Composite Scales over the AFQT for Predicting IMT Physical Fitness Criteria by Education Tier	35
Table 5.5. Incremental Validity Estimates for the TAPAS and TOPS Composite Scales over the AFQT for Predicting IMT Peer Support Criteria by Education Tier	36
Table 5.6. Incremental Validity Estimates for the TAPAS and TOPS Composite Scales over the AFQT for Predicting IMT Commitment and Satisfaction with the Army by Education Tier	37

Table 5.7. Incremental Validity Estimates for the TAPAS and TOPS Composite Scales over the AFQT for Predicting Cumulative Attrition through 18 Months of Service by Education Tier.....	39
Table 5.8. Incremental Validity Estimates for the TAPAS and TOPS Composite Scales over the AFQT for Predicting In-Unit Technical Performance Criteria by Education Tier.....	40
Table 5.9. Incremental Validity Estimates for the TAPAS and TOPS Composite Scales over the AFQT for Predicting In-Unit Effort and Disciplinary Incidents Criteria by Education Tier	41
Table 5.10. Incremental Validity Estimates for the TAPAS and TOPS Composite Scales over the AFQT for Predicting In-Unit Physical Fitness Criteria by Education Tier.....	42
Table 5.11. Incremental Validity Estimates for the TAPAS and TOPS Composite Scales over the AFQT for Predicting In-Unit Peer Support and Leadership Potential Criteria by Education Tier	43
Table 5.12. Incremental Validity Estimates for the TAPAS and TOPS Composite Scales over the AFQT for Predicting In-Unit Commitment and Satisfaction with the Army by Education Tier	44
Table 5.13. Summary of the Bivariate Correlations between AFQT, TAPAS, and Selected IMT Criteria for Tier 1 Soldiers	46
Table 5.14. Summary of the Bivariate Correlations between AFQT, TAPAS, and Selected IMT Criteria and Attrition for Tier 2 Soldiers	47
Table 5.15. Bivariate Correlations between AFQT, TAPAS, and Selected IMT Criteria, and Attrition for Tier 1 + Tier 2 Soldiers.....	48
Table 6.1. TAPAS Dimensions Assessed	50

LIST OF FIGURES

Figure 1.1. TOPS Initial Operational Test & Evaluation (IOT&E).....	3
Figure 2.1. Overview of TOPS data file merging and nested sample generation process.....	5
Figure 4.1. IMT Army-Wide Performance Rating Scale dimensions and composite score composition.....	17
Figure 4.2. Sample IMT 7-point behaviorally-anchored rating scale.	17
Figure 4.3. IMT relative overall performance rating scale.	18
Figure 4.4. In-Unit Army-Wide Performance Rating Scale dimensions and composite score composition.	18

TIER ONE PERFORMANCE SCREEN INITIAL OPERATIONAL TEST AND EVALUATION: 2011 ANNUAL REPORT

CHAPTER 1: INTRODUCTION

Deirdre J. Knapp (HumRRO), Tonia S. Heffner, Leonard A. White, and Kate LaPort (ARI)

Background

The Personnel Assessment Research Unit (PARU) of the U.S. Army Research Institute for the Behavioral and Social Sciences (ARI) is responsible for conducting personnel research for the Army. The focus of PARU's research is maximizing the potential of the individual Soldier through effective selection, classification, and retention strategies.

In addition to educational, physical, and moral screens, the U.S. Army relies on a composite score from the Armed Services Vocational Aptitude Battery (ASVAB), the Armed Forces Qualification Test (AFQT), to select new Soldiers into the Army. Although the AFQT has proven to be, and will continue to serve as, a useful metric for selecting new Soldiers, other personal attributes, in particular non-cognitive attributes (e.g., temperament, interests, and values), are important to entry-level Soldier performance and retention (e.g., Knapp & Tremble, 2007).

In December 2006, the Department of Defense (DoD) ASVAB review panel—a panel of experts in the measurement of human characteristics and performance—released their recommendations (Dragow, Embretson, Kyllonen, & Schmitt, 2006). Several of these recommendations focused on supplementing the ASVAB with additional measures for use in selection and classification decisions. The ASVAB review panel further recommended that the use of these measures be validated against performance criteria.

Just prior to release of the ASVAB review panel's findings, ARI had initiated a longitudinal research effort, *Validating Future Force Performance Measures (Army Class)*, to examine the prediction potential of several non-cognitive measures (e.g., temperament and person-environment fit) for Army outcomes (e.g., performance, attitudes, attrition). The Army Class research project is a 6-year effort that is being conducted with contract support from the Human Resources Research Organization ([HumRRO]; Ingerick, Diaz, & Putka, 2009; Knapp & Heffner, 2009). Experimental predictors were administered to new Soldiers in 2007 and early 2008. Since then, Army Class researchers have obtained attrition data from Army records and collected training criterion data on a subset of the Soldier sample. Job performance criterion data were collected from Soldiers in the Army Class longitudinal validation sample in 2009 (Knapp, Owens, & Allen, 2011). A second round of job performance data collection was completed in April 2011. Final analysis and reporting of this program of research is nearing completion.

After the Army Class research was underway, ARI initiated the *Expanded Enlistment Eligibility Metrics (EEEM)* project (Knapp & Heffner, 2010). The EEEM goals were similar to Army Class, but the focus was specifically on Soldier selection and the time horizon was much shorter. Specifically, EEEM required identification of one or more promising new predictor measures for immediate implementation. The EEEM project capitalized on the existing Army Class data collection procedure and, thus, the EEEM sample was a subset of the Army Class sample.

As a result of the EEEM findings, Army policy-makers approved an initial operational test and evaluation (IOT&E) of the *Tier One Performance Screen (TOPS)*. This report is the fourth in a series presenting continuing analyses from the IOT&E of TOPS.

The Tier One Performance Screen (TOPS)

Six experimental pre-enlistment measures were included in the EEEM research (Allen, Cheng, Putka, Hunter, & White, 2010). These included several temperament measures, a situational judgment test, and two person-environment fit measures based on values and interests. The most promising measures recommended to the Army for implementation were identified based on the following considerations:

- Incremental validity over AFQT for predicting important performance and retention-related outcomes
- Minimal subgroup differences
- Low susceptibility to response distortion (e.g., faking good)
- Minimal administration time requirements

The Tailored Adaptive Personality Assessment System ([TAPAS]; Stark, Chernyshenko, & Drasgow, 2010b) surfaced as the top choice, with the Work Preferences Assessment ([WPA]; Putka & Van Iddekinge, 2007) identified as another good option that was substantively different from the TAPAS. Specifically, TAPAS is a measure of personality characteristics (e.g., achievement, sociability) that capitalizes on the latest advances in psychometric theory and provides a good indicator of personal motivation. The WPA asks applicants to indicate their preference for various kinds of work activities and environments (e.g., “A job that requires me to teach others,” “A job that requires me to work outdoors”). Although not included in the EEEM research, the Information/Communications Technology Literacy (ICTL) test emerged as a potential test of applicants’ familiarity with computers and information technology, which may predict performance in high-technology occupations.

In May 2009, the Military Entrance Processing Command (MEPCOM) began administering TAPAS on the computer adaptive platform for the ASVAB (CAT-ASVAB). Initially, TAPAS was to be administered only to Education Tier 1, non-prior service applicants.¹ The limitation to Tier 1 was removed early in CY2011 so the Army could evaluate TAPAS across all types of applicants.

TOPS uses non-cognitive measures to identify applicants who would likely perform differently (higher or lower) than would be predicted by their ASVAB scores. As part of the TOPS IOT&E, TAPAS scores are being used to screen out a small number of AFQT Category IIIB/ IV applicants.² Although the WPA is part of the TOPS IOT&E, WPA scores will not be considered for enlistment eligibility. The WPA is being prepared for MEPS administration starting in CY2012.

¹ Applicant educational credentials are classified as Tier 1 (primarily high school diploma), Tier 2 (primarily non-diploma graduate), and Tier 3 (not a high school graduate).

² Examinees are classified into categories based on their AFQT percentile scores (Category I = 93-99, Category II = 65-92, Category IIIA = 50-64, Category IIIB = 31-49, Category IV = 10-30, Category V = 1-9).

Although the initial conceptualization for the IOT&E was to use TAPAS as a tool for “screening in” Education Tier 1 applicants with lower AFQT scores, changing economic conditions spurred a reconceptualization to a system that screens out low motivated applicants across AFQT score categories. It is likely that the selection model in a fully operational system would adjust to fit with the applicant market that changes depending on the economy and other factors.

Evaluating TOPS

Figure 1.1 illustrates the TOPS IOT&E research plan. To evaluate the non-cognitive measures (TAPAS and WPA), the Army is collecting training criterion data on Soldiers in eight target military occupational specialties (MOS) as they complete initial military training (IMT).³ The criterion measures include job knowledge tests (JKTs); an attitudinal assessment, the Army Life Questionnaire (ALQ); and performance rating scales (PRS) completed by the Soldiers’ cadre. These measures are computer-administered at the schools for each of the eight target MOS. The process is overseen by Army personnel with guidance and support from both ARI and HumRRO. Course grades and completion rates are obtained from administrative records for all Soldiers who take the TAPAS, regardless of MOS.

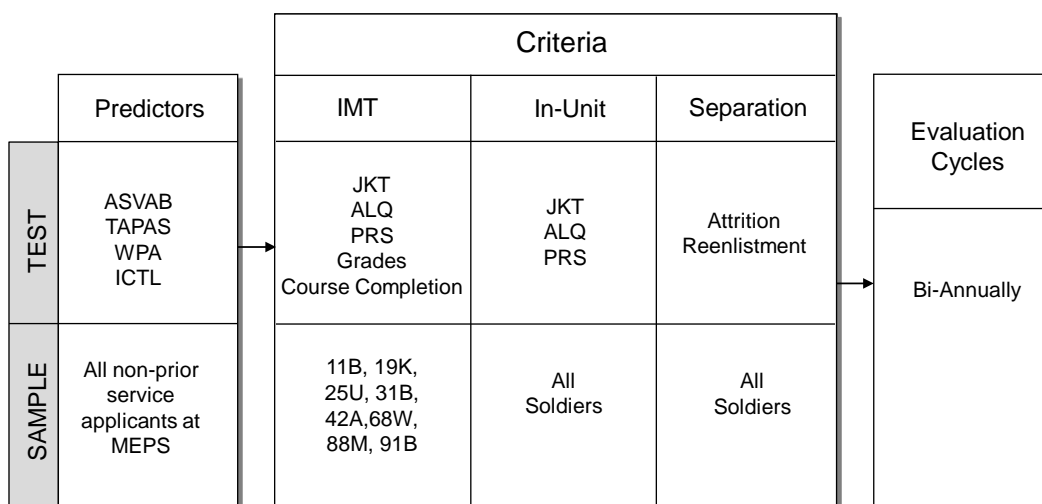


Figure 1.1. TOPS Initial Operational Test & Evaluation (IOT&E).

Criterion data were also collected from Soldiers in their units. This “in-unit” data collection began in January 2001 and targeted Soldiers who took the TAPAS prior to enlistment. The in-unit criterion measures include JKTs, the ALQ, and supervisor ratings of performance. The data collection model closely mirrors that which was used in the Army Class research program (Knapp, Owens, et al., 2011). We are visiting major Army installations and have plans to visit

³ The target MOS are Infantryman (11B), Armor Crewman (19K), Signal Support Specialist (25U), Military Police (31B), Human Resources Specialist (42A), Health Care Specialist (68W), Motor Transport Operator (88M), and Light Wheel Vehicle Mechanic (91B). These MOS were selected to include large, highly critical MOS as well as to represent the diversity of work requirements across MOS.

reserve component training sites to collect Soldier and supervisor data in proctored settings. Separation status of all Soldiers who took the TAPAS prior to enlistment is being tracked throughout the course of the research.

This report describes the fourth iteration to develop a criterion-related validation data file and conduct evaluation analyses using data collected in the TOPS IOT&E initiative. Prior evaluations are described in Knapp, Heffner, and White (2011) and Knapp and Heffner (2011, 2012). Additional analysis datasets and validation analyses will be prepared and conducted at 6-month intervals throughout the multi-year IOT&E period.

Overview of Report

Chapter 2 explains how the evaluation analysis data files are constructed and then describes characteristics of the samples resulting from construction of the latest analysis data file. Chapter 3 describes the TAPAS and ASVAB, including content, scoring, and psychometric characteristics. Chapter 4 describes the training and in-unit criterion measures included in this analysis, including their psychometric characteristics. Note that this is the first evaluation cycle in which we have been able to include in-unit criterion data. Criterion-related validation analyses are presented in Chapter 5. The report concludes with Chapter 6, which summarizes our continuing efforts to evaluate TOPS and looks toward plans for future iterations of these evaluations.

CHAPTER 2: DATA FILE DEVELOPMENT

D. Matthew Trippe, Laura Ford, Bethany Bynum, and Karen Moriarty (HumRRO)

Overview of Process

The TOPS data file is assembled from a number of sources. In general, the data file comprises predictor and criterion data obtained from administrative, IMT (or “schoolhouse”), and in-unit sources. The IMT and in-unit assessments are described in Chapter 4.

An illustrative view of the TOPS analysis file construction process is provided in Figure 2.1.⁴ The lighter boxes within the figure represent source data files, and the darker boxes represent samples on which descriptive or inferential analyses are conducted. Samples are formed by applying filters to a data file such that it includes the observations of interest. The leftmost column in the figure summarizes the predictor data sources used to derive the TOPS Applicant Sample. The other columns summarize the research-only (i.e., non-administrative) and

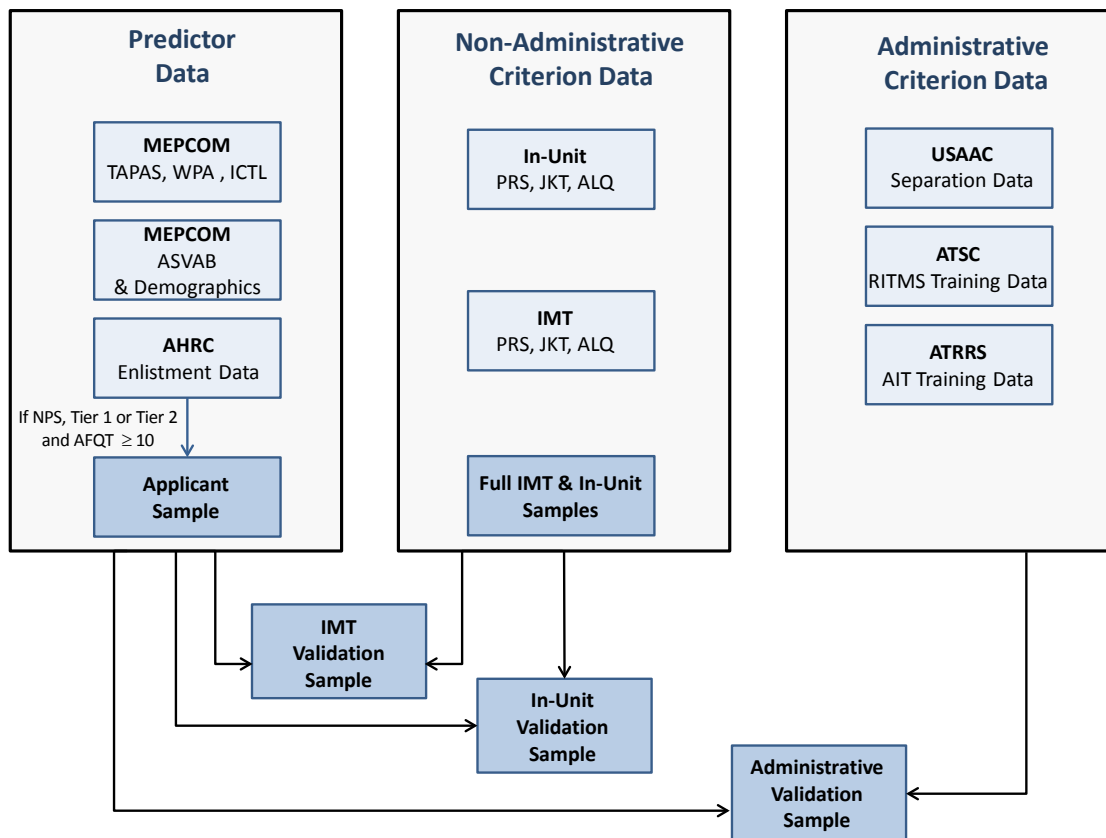


Figure 2.1. Overview of TOPS data file merging and nested sample generation process.

⁴ Administrative data are collected from the following sources: (a) Military Entrance Processing Command (MEPCOM), (b) Army Human Resources Command (AHRC), (c) U.S. Army Accessions Command (USAAC), (d) Army Training Support Center’s (ATSC) Resident Individual Training Management System (RITMS), and (e) Training and Doctrine Command’s (TRADOC) Army Training Requirements and Resources System (ATRRS).

administrative criterion data. Predictor and criterion data are merged to form the IMT/in-unit-specific validation samples and the large administrative validation sample, which include all Soldiers who have predictor data and at least one criterion record (e.g., administrative data). The latest version of the TOPS data file does not contain WPA predictor scores or a sufficient number of ICTL scores to warrant inclusion in the analysis data set at this time. Future versions of the data file will include those data.

Description of Data File and Sample Construction

The latest data file created in January 2012 includes TAPAS data collected through July 2011 (after which new TAPAS forms were introduced) and criterion data collected through December 2011. Table 2.1 summarizes the relevant characteristics of the total TAPAS sample contained in the January 2012 TOPS data file. The total sample includes applicants who did not enlist in the Army. The TOPS Applicant Sample was defined by limiting records in the total sample data file provided by MEPCOM to those Soldiers who are non-prior service, Education Tier 1 or 2, have an AFQT score of 10 or greater, and have a pre-August 2011 TAPAS score. New versions of the TAPAS were introduced in August 2011 and there are insufficient numbers of cases to include in the evaluation work at this time. Among the 216,565 applicants in the total, unfiltered sample, 176,467 (81.5%) met these screens and constituted the Applicant Sample that served as the basis for the analyses summarized in this report. Sample sizes reported in subsequent chapters will generally be smaller than the initial numbers reported here because of further data filtering or disaggregation that occurs for each particular analysis. Predictor and criterion scores were determined to be valid if they passed multiple data quality screens intended to identify unmotivated applicants. Those additional screens have not yet been applied to the samples described in this chapter because they are often specific to a particular analysis.

A detailed breakout of background and demographic characteristics observed in the analytic samples appears in Table 2.2. Regular Army Soldiers comprise a majority of the cases in each sample. AFQT categories follow an expected distribution. The samples are predominantly male, Caucasian, and non-Hispanic; however, a large percentage of Soldiers declined to provide information on race or ethnicity.

The Administrative Validation Sample described in Table 2.2 includes 69,495 Soldiers. Included in this sample are Soldiers who meet all of the inclusion criteria for the TOPS Applicant Sample and also have at least one record in a criterion data source (i.e., Army Training Requirements and Resources System [ATTRS], Resident Individual Training Management System [RITMS], IMT, in-unit, attrition). However, the number of Soldiers included in any individual analysis is generally much smaller. The exact number of Soldiers varies by criterion depending on the availability of valid data on key variables. Specific sample details on each criterion variable are provided in subsequent chapters.

Although there are 36,401 Soldiers in the Full IMT data file, only 9,502 had taken the TAPAS when they applied for enlistment. This disconnect is largely due to the fact that most of the Soldiers tested at the schools had taken their pre-enlistment tests before MEPCOM started administering the TAPAS widely to applicants. The problem is exacerbated by the gradual introduction of the TAPAS across MEPS locations so that early in the IOT&E, not all MEPS

were actively participating. Another contributing factor is the extended time, ranging from approximately 6-9 months, from when the applicants complete enlistment testing and their access into the Army. We expect that future analysis data files will continue to show a higher match between Soldiers tested in the schools and those tested pre-enlistment. Indeed, the match rate at this stage (26.8%) is an improvement over the match rates obtained previously (19.8%, 12.7%, 5.5%; Trippe, Ford, Bynum, & Moriarty, 2012). Similarly, there are 1,861 Soldiers with in-unit data, but only 387 ultimately matched to a TAPAS record. There are 80 Soldiers with a TAPAS record and both IMT and in-unit criterion data.

Table 2.1. Full TAPAS Data File Characteristics

Variables	<i>n</i>	% of Total Sample (<i>N</i> = 216,565)
<i>Education Tier</i>		
Tier 1	201,214	92.9
Tier 2	10,748	5.0
Tier 3	4,598	2.1
<i>Prior Service</i>		
Yes	5,319	2.5
No or Missing	211,246	97.5
<i>Military Occupational Specialty</i>		
11B/11C/11X/18X	16,025	7.4
19K	943	0.4
25U	1,754	0.8
31B	3,589	1.6
42A	1,865	0.9
68W	4,824	2.2
88M	4,902	2.3
91B	4,093	1.9
Other	55,169	25.5
Unknown ^a	123,401	57.0
<i>AFQT Category</i>		
I	15,583	7.2
II	65,098	30.1
IIIA	42,534	19.6
IIIB ^b	59,315	27.4
IV ^b	30,731	14.2
V	3,296	1.5
<i>Contract Status</i>		
Signed	124,703	57.6
Not signed	91,862	42.4
Applicant Sample ^c	176,467	81.5

^a Generally, when the MOS is unknown, it is either because the respondent did not access into the Army or because the information was not yet available in the data sources on which the January 2012 data file was based.

^b AFQT Categories IIIB and IV are oversampled. Figures presented are not representative of Army accessions.

^c The Applicant Sample size is smaller than the total TAPAS sample because it is limited to non-prior service, Education Tier 1 and 2, AFQT ≥ 10 applicants with pre-August 2011 TAPAS scores.

Table 2.2. Background and Demographic Characteristics of the TOPS Samples

Characteristic	Applicant ^a <i>n</i> = 176,467		Administrative Validation ^b <i>n</i> = 69,495		IMT Validation ^c <i>n</i> = 9,502		In-Unit Validation ^d <i>n</i> = 387	
	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%
<i>Component</i>								
Regular	108,657	61.6	41,292	59.4	5,834	61.4	380	98.2
ARNG	45,416	25.7	19,471	28.0	2,856	30.1	--	--
USAR	22,251	12.6	8,727	12.6	812	8.6	--	--
<i>Education Tier</i>								
Tier 1	168,189	95.3	67,489	97.1	9,221	97.0	378	97.7
Tier 2	8,278	4.7	2,006	2.9	281	3.0	9	2.3
<i>Military Occupational Specialty</i>								
11B/11C/11X/18X	14,428	8.2	12,656	18.2	4,128	43.4	91	23.5
19K	874	0.5	796	1.2	100	1.1	9	2.3
25U	1,552	0.9	1,166	1.7	10	0.1	5	1.3
31B	3,038	1.7	2,417	3.5	1,561	16.4	8	2.1
42A	1,578	0.9	1,266	1.8	108	1.1	5	1.3
68W	4,414	2.5	3,828	5.5	1,833	19.3	19	4.9
88M	4,185	2.4	3,253	4.7	1,421	15.0	18	4.6
91B	3,556	2.0	2,841	4.1	339	3.6	18	4.6
Other	49,314	26.1	41,213	59.3	--	--	214	55.3
Unknown	93,528	53.0	68	0.1	--	--	--	--
<i>AFQT Category</i>								
I	13,058	7.4	5,908	8.5	804	8.5	37	9.6
II	54,809	31.1	25,226	36.3	3,819	40.2	117	30.2
IIIA	35,564	20.2	15,764	22.7	2,056	21.7	89	23.0
IIIB	47,516	26.9	19,030	27.4	2,424	25.5	133	34.4
IV	25,520	14.5	3,567	5.1	399	4.2	11	2.8
<i>Gender</i>								
Female	34,755	19.7	11,531	16.6	1,221	12.9	50	12.9
Male	141,056	79.9	57,900	83.3	8,277	87.1	337	87.1
<i>Race</i>								
African American	28,680	16.3	9,397	13.5	960	10.1	47	12.1
American Indian	1,292	0.7	487	0.7	78	0.8	1	0.3
Asian	5,460	3.1	2,248	3.2	272	2.9	3.1	3.1
Hawaiian/Pacific Islander	1,231	0.7	532	0.8	75	0.8	2	0.5
Caucasian	126,738	71.8	53,513	77.0	7,616	80.2	306	79.1
Multiple	763	0.4	301	0.4	48	0.5	--	--
Declined to Answer	12,284	7.0	3,015	4.3	453	4.8	19	4.9
<i>Ethnicity</i>								
Hispanic/Latino	25,778	14.6	9,617	13.8	1,108	11.6	44	11.4
Not Hispanic	138,721	78.6	57,308	82.5	8,009	84.3	329	85.0
Declined to Answer	11,948	6.8	2,569	3.7	385	4.0	14	3.6

Note. Values in the % columns may not sum to 100 because of missing data.

^a Limited to applicants who had no prior service, Education Tier 1 or 2, AFQT ≥ 10 , and a pre-August 2011 TAPAS score; served as the core analysis sample.

^b Soldiers in Applicant Sample with at least one criterion record (i.e., schoolhouse, in-unit, ATTRS, RITMS, or attrition).

^c Soldiers in Applicant Sample with criterion data collected at schoolhouses.

^d Soldiers in Applicant Sample with criterion data collected from Soldiers in units.

Summary

The TOPS data file is periodically updated by merging new TAPAS scores, administrative records, IMT, and in-unit data into one master data file. The January 2012 data file includes a total of 216,565 applicants who took the TAPAS, 176,467 of whom were in the TOPS Applicant Sample. The Applicant Sample was determined by excluding Education Tier 3, AFQT Category V, and prior service applicants from the master data file. Of that Applicant Sample, 69,495 (39.4%) had a record in at least one of the administrative criterion data sources; 9,502 had IMT data collected from the schoolhouse and 387 had in-unit criterion data. The IMT match rate represents an improvement from the prior reporting cycle. This is likely due to the maturation of criterion data in the source data files. Higher match rates observed in the present reporting cycle are likely to improve the stability and interpretability of results over the prior cycle. Nevertheless, the amount of criterion data that is actually used in a given analysis remains small in relation to the amount of available predictor data. Subsequent iterations of the TOPS IOT&E data file will no doubt show progressively stronger sample sizes to support validation and other evaluative analyses.

CHAPTER 3: DESCRIPTION OF THE TOPS IOT&E PREDICTOR MEASURES

Stephen Stark, O. Sasha Chernyshenko, Fritz Drasgow (Drasgow Consulting Group), and Matthew T. Allen (HumRRO)

The purpose of this chapter is to describe the predictor measures investigated to date in the TOPS IOT&E (i.e., TAPAS and ASVAB). The central predictor under investigation in this analysis is TAPAS (Stark, Chernyshenko, & Drasgow, 2010b), while the baseline predictor used by the Army is the ASVAB. Two additional experimental measures, the ICTL and WPA, are not yet included in the analysis data files and are therefore not discussed further here.

Tailored Adaptive Personality Assessment System (TAPAS)

Description

TAPAS is a personality measurement tool developed by Drasgow Consulting Group (DCG) under the Army's Small Business Innovation Research (SBIR) program. The system builds on the foundational work of the Assessment of Individual Motivation ([AIM]; White & Young, 1998) by incorporating features designed to promote resistance to faking and by measuring narrow personality constructs (i.e., facets) that are known to predict outcomes in work settings. Because TAPAS uses item response theory (IRT) methods to construct and score items, it can be administered in multiple formats: (a) as a fixed length, *non-adaptive test* where examinees respond to the same sequence of items or (b) as an *adaptive test* where each examinee responds to a unique sequence of items selected to maximize measurement accuracy for that specific examinee.

TAPAS uses an IRT model for multidimensional pairwise preference items ([MUPP]; Stark, Chernyshenko, & Drasgow, 2005) as the basis for constructing, administering, and scoring personality tests that are designed to reduce response distortion (i.e., faking) and yield normative scores even with tests of high dimensionality (Stark, Chernyshenko, & Drasgow 2010a). TAPAS items consist of pairs of personality statements for which a respondent's task is to choose the one that is "more like me." The two statements constituting each item are matched in terms of social desirability and often represent different dimensions. As a result, respondents have a difficult time discerning which answers improve their chances of being enlistment eligible. Because they are less likely to know which dimensions are being used for selection, they are less likely to discern which statements measure those dimensions, and they are less likely to be able to keep track of their answers on several dimensions simultaneously so as to provide consistent patterns of responses across the whole test. Without knowing which answers have an impact on their eligibility status, respondents should not be able to increase their scores on selection dimensions as easily as when traditional, single statement measures are used.

The use of a formal IRT model also greatly increases the flexibility of the assessment process. A variety of test versions can be constructed to measure personality dimensions that are relevant to specific work contexts, and the measures can be administered via paper-and-pencil or computerized formats. If test content specifications (i.e., test blueprints) are comparable across versions, the respective scores can be readily compared because the metric of the statement

parameters has already been established by calibrating response data obtained from a base or reference group (e.g., Army recruits). The same principle applies to adaptive testing, wherein each examinee receives a different set of items chosen specifically to reduce the error in his or her trait scores at points throughout the exam. Adaptive item selection enhances test security because there is less overlap across examinees in terms of the items presented.

Another important feature of TAPAS is that pools of statements representing 22 narrow personality traits are available. The TAPAS trait taxonomy was developed using the results of several large scale factor-analytic studies with the goal of identifying a comprehensive set of non-redundant narrow traits. These narrow traits, if necessary or desired, can be combined to form either the Big Five (the most common organization scheme for narrow personality traits) or any other number of broader traits (e.g., Integrity or Positive Core Self-Evaluations). This is advantageous for applied purposes because TAPAS versions can be created to fit a wide range of applications and are not limited to a particular service branch or criterion. Selection of specific TAPAS dimensions can be guided by consulting the results of a meta-analytic study performed by DCG that mapped the 22 TAPAS dimensions to several important organizational criteria for military and civilian jobs (e.g., task proficiency, training performance, attrition) (Chernyshenko & Stark, 2007).

Scoring details and the criterion-related validation work that led to its inclusion in the TOPS IOT&E can be found in the *Expanded Enlistment Eligibility Metrics* report (Knapp & Heffner, 2010) and in prior evaluation reports in this series. For the analyses in this report, cases were dropped based on incomplete response data and evidence of unmotivated responding.

Multiple Versions of TAPAS

As part of the TOPS IOT&E, multiple versions of the TAPAS have been administered as ARI explores the value of new and alternative dimensions. In the present report, results are based on two 15-dimension versions of TAPAS, each administering 120 items (i.e., pairs of statements). One version was nonadaptive (static), so all examinees answered the same sequence of items; the other was adaptive, so each examinee answered items tailored to his or her trait level estimates. The TAPAS-15D-Static was administered from mid-July to mid-September of 2009 to all examinees, and later to smaller numbers of examinees at some MEPS. The adaptive version, referred to as TAPAS-15D-CAT, was introduced in September 2009. These 15-dimension versions were replaced by three 21-dimension versions that have been administered since August 2011. These new versions will be described more fully in the next evaluation report where we expect to have sufficient criterion data to include them in the validation analyses. Table 3.1 shows the facets assessed by the 15-dimension measures.

Scores across TAPAS versions are combined by:

1. Filtering out participants who were not part of the sample of interest (i.e., those that were not in the “TOPS Applicant Sample”—Education Tier 1 and 2, non-prior service, AFQT Category IV or above), and
2. Standardizing the variables within version using a z -transformation, computed by subtracting each score from the mean for that version and dividing by the standard deviation.

Table 3.1. TAPAS Dimensions Assessed

Facet Name	Brief Description	“Big Five” Broad Factor
Dominance	High scoring individuals are domineering, “take charge” and are often referred to by their peers as "natural leaders."	Extraversion
Sociability	High scoring individuals tend to seek out and initiate social interactions.	
Attention Seeking	High scoring individuals tend to engage in behaviors that attract social attention; they are loud, loquacious, entertaining, and even boastful.	
Generosity	High scoring individuals are generous with their time and resources.	Agreeableness
Cooperation	High scoring individuals are trusting, cordial, non-critical, and easy to get along with.	
Achievement	High scoring individuals are seen as hard working, ambitious, confident, and resourceful.	Conscientiousness
Order	High scoring individuals tend to organize tasks and activities and desire to maintain neat and clean surroundings.	
Self Control	High scoring individuals tend to be cautious, levelheaded, able to delay gratification, and patient.	
Non-Delinquency	High scoring individuals tend to comply with rules, customs, norms, and expectations, and they tend not to challenge authority.	
Adjustment	High scoring individuals are worry free, and handle stress well; low scoring individuals are generally high strung, self-conscious and apprehensive.	Emotional Stability
Even Tempered	High scoring individuals tend to be calm and stable. They don’t often exhibit anger, hostility, or aggression.	
Optimism	High scoring individuals have a positive outlook on life and tend to experience joy and a sense of well-being.	
Intellectual Efficiency	High scoring individuals are able to process information quickly and would be described by others as knowledgeable, astute, and intellectual.	Openness To Experience
Tolerance	High scoring individuals are interested in other cultures and opinions that may differ from their own. They are willing to adapt to novel environments and situations.	
Physical Conditioning	High scoring individuals routinely participate in vigorous sports or exercise and enjoy physical work.	Other

TAPAS Composites

In addition to the validation analyses described above, an initial Education Tier 1 performance screen was developed from the TAPAS-95s scales for the purpose of testing in an applicant setting (Allen et al., 2010).⁵ This was accomplished by (a) identifying key criteria of most interest to the Army, (b) sorting these criteria into “can-do” and “will-do” categories (see below), and (c) selecting composite scales corresponding to the can-do and will-do criteria, taking into account both theoretical rationale and empirical results. The result of this process was two composite scores.

1. **Can-Do Composite**: The TOPS Can-Do composite consists of five TAPAS scales and is designed to predict the extent to which Soldiers can perform the technical aspects of their jobs, using indicators such as MOS-specific job knowledge, Advanced Individual Training (AIT) exam grades, and graduation from AIT/One Station Unit Training (OSUT).
2. **Will-Do Composite**: The TOPS Will-Do composite consists of five TAPAS scales (three of which overlap with the Can-Do composite) and is designed to predict the more motivational elements of job performance, such as maintaining physical fitness, adjusting to Army life, demonstrating effort, and supporting peers.

The analyses on which these composites were based focused on Tier 1 AFQT Category IIIB applicants. Due to changing recruitment priorities (as described in Chapter 1), the initial target group for the TOPS IOT&E was AFQT Category IV applicants. Subsequently, the TOPS IOT&E was expanded to include all Tier 1 and Tier 2 applicants above AFQT Category V, but screening based on TAPAS scores is confined to Category IIIB and IV Tier 1 applicants.

As more data become available and the dimensions included in the different TAPAS versions stabilizes, new TAPAS composites will be developed and evaluated.

Armed Services Vocational Aptitude Battery (ASVAB) Content, Structure, and Scoring

The ASVAB is a multiple aptitude battery of nine tests administered by the MEPCOM. Most military applicants take the computer adaptive version of ASVAB (i.e., the CAT-ASVAB). Scores on the ASVAB tests are combined to create composite scores for use in (a) selecting applicants into the Army and (b) classifying them to an MOS. The AFQT, the composite used for selecting applicants into the Army, comprises the Verbal Expression⁶ (VE), Arithmetic Reasoning (AR), and Math Knowledge (MK) tests ($AFQT = 2*VE + AR + MK$). Applicants must meet a minimum AFQT score to be eligible to serve in the military, and the Services favor

⁵ TAPAS-95s was a paper-and-pencil, static version of the TAPAS used in the Army Class research.

⁶ Verbal Expression is a scaled combination of the Word Knowledge (WK) and Paragraph Comprehension (PC) tests.

high-scoring applicants for enlistment (e.g., through enlistment bonuses). AFQT percentile scores are divided into the following categories:⁷

- Category I (93-99)
- Category II (65-92)
- Category IIIA (50-64)
- Category IIIB (31-49)
- Category IV (10-30)
- Category V (1-9)

AFQT Category V Soldiers are not eligible for enlistment, Category IV accessions are greatly restricted, and priority is given to Category I-IIIa accessions.

For classification, scores on the ASVAB tests are combined to form nine Aptitude Area (AA) composites.⁸ An applicant must receive a minimum score on the MOS-relevant AA composite(s) to qualify for classification to that MOS. For example, applicants must score a 95 in both the Electronics (EL) and Signal Communications (SC) AA composites to qualify as a Signal Support Specialist (25U). Descriptive statistics for the AFQT, ASVAB tests, and AA composites are reported in Table A.5 in Appendix A. AFQT Category frequencies are reported in Chapter 2 (Tables 2.2 and 2.3).

Summary

The purpose of this chapter was to describe the predictor measures used as part of the TOPS IOT&E. The TAPAS is unique among personality measures because it uses forced-choice pairwise items and IRT to promote resistance to faking. Initial validation research conducted as part of EEEM was promising enough to warrant an IOT&E. The ASVAB, which consists of multiple tests that are formed into selection (i.e., AFQT) and classification (i.e., AA) composites, is used as the baseline instrument for incremental validity analyses reported in Chapter 5.

⁷ For more information on ASVAB scoring, see the official website of the ASVAB, www.officialasvab.com.

⁸ A tenth AA composite, General Technical (GT), is not used for entry-level enlisted Army selection so is not included here.

CHAPTER 4: DESCRIPTION AND PSYCHOMETRIC PROPERTIES OF CRITERION MEASURES

Karen O. Moriarty, Thomas Kiger, and Chad Peddie (HumRRO)

Criterion measures including job knowledge tests (JKTs), performance rating scales (PRS), and attitudinal data captured on a self-report questionnaire (ALQ) were used to validate the TAPAS. These measures were originally developed for the Army Class project (Moriarty, Campbell, Heffner, & Knapp, 2009), and modified, where needed, for inclusion in the TOPS IOT&E. These measures were used to supplement administrative data available from Army records. Table 4.1 summarizes the criterion measures, noting differences between the instruments used to collect from Soldiers in IMT (training) and Soldiers in units.

In this chapter, we first describe the criterion measures in more detail, including revisions that were made during the course of the data collection. The chapter concludes with discussion of the distributional and psychometric characteristics of the various measures.

Table 4.1. Summary of IMT and In-Unit Criterion Measures

Criterion Measure	Description
<i>Soldier/Cadre Reported</i>	
Job Knowledge Tests (JKT)	The Warrior Tasks and Battle Drills (WTBD) JKT measures knowledge that is general to all enlisted Soldiers. MOS-specific JKTs measure Soldiers' knowledge of basic facts, principles, and procedures required of Soldiers in training for a particular MOS. Each JKT includes a mix of item formats (e.g., multiple-choice, multiple-response, and rank order).
Performance Rating Scales (PRS)	Training PRS measure Soldiers' performance in two categories: (a) MOS-specific (e.g., learns preventive maintenance checks and services, learns to troubleshoot vehicle and equipment problems) and (b) Army-wide (e.g., exhibits effort, supports peers, demonstrates physical fitness). The IMT PRS are completed by drill sergeants or training cadre. In-unit PRS cover Army-wide dimensions only and are completed by supervisors.
Army Life Questionnaire (ALQ)	The ALQ measures Soldiers' self-reported attitudes and experiences in the Army. The IMT and in-unit versions are very similar.
<i>Administrative</i>	
Attrition	Separation data are obtained on participating Regular Army Soldiers at 3 months (attrition near or after the completion of Basic Combat Training), 4 months (attrition during AIT/OSUT), 6 months (attrition near or after completion of AIT/OSUT), and at regular 3-month intervals thereafter. Attrition data were available out to 18 months for the current sample.
Initial Military Training (IMT) Criteria	These data provide information concerning how many Soldiers restarted IMT and for what reasons, the number of times Soldiers restarted training, and graduation status.
School Grades	Schoolhouse grades for Soldiers in Advanced Individual Training (AIT).

Criterion Measure Descriptions

Job Knowledge Tests (JKTs)

Multiple sets of JKTs (IMT and in-unit) were developed or adapted from the Select21 (Collins, Le, & Schantz, 2005) and Army Class (Moriarty et al., 2009) projects: one for Warrior Tasks and Battle Drills (WTBD), which is administered to all participating Soldiers, and a set of MOS-specific JKTs for Infantry, Armor, Military Police, Health Care Specialist, Light Wheel Vehicle Mechanic, and Motor Transport Operator Soldiers. The Army Class research did not include two MOS targeted in the TOPS IOT&E— Signal Support Specialist (25U) and Human Resources Specialist (42A). MOS-specific JKTs for these two MOS have recently been developed, though there are not yet sufficient data to include them in the analysis data file.

Most of the JKT items are in a multiple-choice format with two to four response options. However, other formats, such as multiple-response (i.e., check all that apply), rank ordering, and matching are also used. The items use visual images to make them more realistic and reduce reading requirements for the test.

Prior to finalizing the items for initial use in the TOPS IOT&E, the items were reviewed by project staff and Army SMEs to ensure they were of high quality. The JKTs were reviewed again in the summer of 2011. Poorly performing or outdated items were replaced and additional items were included to ensure adequate coverage of content areas identified in the test blueprints that had been established for each test.

Performance Rating Scales (PRS)

The PRS, like the JKTs, also have roots in previous research (see Moriarty et al., 2009 for details) and include both Army-wide and MOS-specific dimensions. The IMT and in-unit PRS are fairly different, so they will be described separately.

IMT PRS

Figure 4.1 lists the eight Army-wide performance dimensions measured by the IMT scales. Several of the individual dimension ratings were combined so that the measure yields five Army-wide dimension-level scores. We also computed an Army-wide composite rating by averaging the ratings on the eight individual dimension ratings. The number of MOS-specific dimensions ranges from five to nine, which were combined into an overall MOS performance score.⁹ The ratings were completed by cadre members (supervisors/trainers) of the target Soldiers. The scales range from 1 (lowest) to 7 (highest) and include a “not observed” option for instances where the cadre did not have an opportunity to observe a Soldier’s performance on a particular dimension.

⁹ MOS-specific dimension ratings are combined into a single overall MOS performance score because: (a) dimension ratings tend to be highly correlated with each other and (b) inter-reliability estimates obtained in prior research for these scales were low.

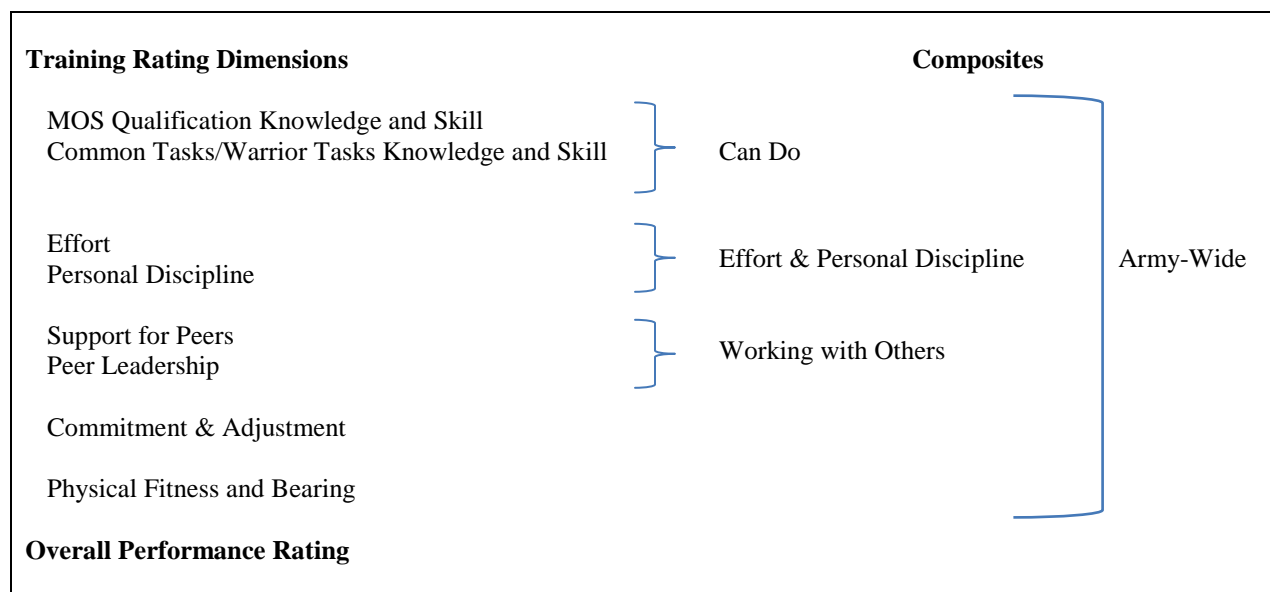


Figure 4.1. IMT Army-Wide Performance Rating Scale dimensions and composite score composition.

The analyses in this report are based on Army-wide and MOS-specific PRS that are in the format of a behaviorally-anchored rating scale (BARS). To assist in their ratings, the scale for each performance dimension includes descriptions (called “anchors”) of high, medium, and low performance. Figure 4.2 provides an example of one of the BARS administered.

In addition to the BARS ratings of each performance dimension, raters were also asked to provide one rating assessing overall performance. This rating was made on a 5-point relative comparison scale, as shown in Figure 4.3. The PRS assessment also includes a 3-point “familiarity” rating in which the rater indicates his or her general opportunity to observe each Soldier being rated (i.e., limited, reasonable, or a lot of opportunity to observe).

Effort						
Puts forth individual effort in study, practice, preparation, and participation activities to complete AIT/OSUT requirements and to meet individual Soldier expectations.						
1	2	3	4	5	6	7
– Puts off studying and practicing tasks.		– Usually completes required assignments.		– Completes study and practice assignments, including non-class requirements, on time.		
– May tune out while an instructor is speaking and sometimes isn't prepared for class.		– Pays attention in class and is usually adequately prepared for class.		– Pays attention in class and studies hard in preparation for class.		
– Tends to give up on tasks if problems arise.		– Usually keeps trying when problems arise.		– Persists with tasks even when problems arise.		

Figure 4.2. Sample IMT 7-point behaviorally-anchored rating scale.

Overall Performance				
Considering your evaluation of the Soldier on the dimensions important to successful performance, please rate the overall effectiveness of each Soldier compared to his/her peers.				
1	2	3	4	5
Among the Weakest	Below Average	Average	Above Average	Among the Best
(in the bottom 20% of Soldiers)	(in the bottom 40% of Soldiers)	(better than the bottom 40% of Soldiers, but not as good as the top 40%)	(in the top 40% of Soldiers)	(in the top 20% of Soldiers)

Figure 4.3. IMT relative overall performance rating scale.

In-Unit PRS

The in-unit PRS uses the 7-point BARS format originally used for the IMT scales (see Figure 4.2). The in-unit Army-wide PRS measured 12 performance dimensions, plus a Leadership Potential scale (see Figure 4.4). A thirteenth scale was dropped for poor psychometric performance and has since been replaced with an Adjustment scale which will be available in future evaluation reports. As with the IMT scales, the individual dimension ratings were combined to form four dimension-level ratings. An Army-wide composite rating was computed by averaging all 12 dimension-level ratings. This is in addition to a single Overall Leadership Potential Rating that is analogous to the IMT Overall Performance Rating.

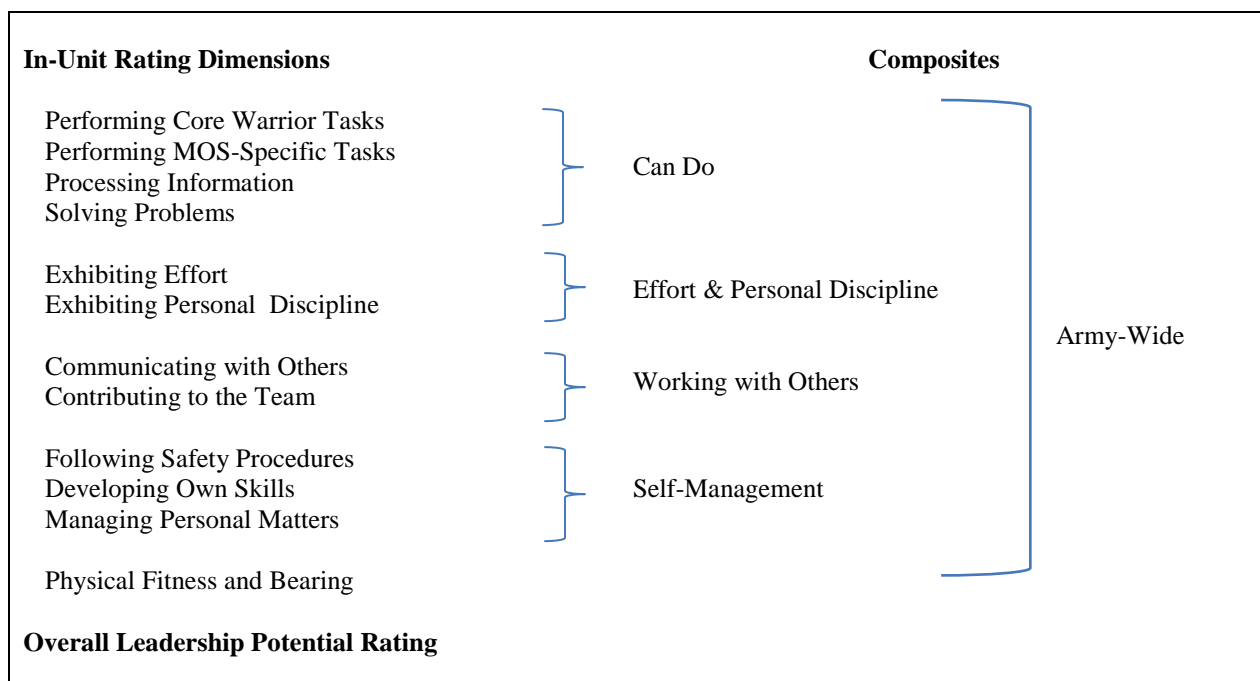


Figure 4.4. In-Unit Army-Wide Performance Rating Scale dimensions and composite score composition.

Army Life Questionnaire (ALQ)

The ALQ was designed to measure Soldiers' self-reported attitudes and experiences in the Army. Earlier forms of the in-unit and training versions of the ALQ (Van Iddekinge, Putka, & Sager, 2005) were modified slightly for use in the TOPS IOT&E. The ALQ includes scales that cover (a) Soldiers' commitment and retention-related attitudes and (b) Soldiers' performance and adjustment. Each ALQ scale is scored differently depending on the nature of the attribute being measured. The Army Physical Fitness Test (APFT) score is a write-in item. Training Achievements, Training Failures, (both of which appear only on the IMT version of the ALQ) and Disciplinary Incidents are simply a sum of the "yes" responses. The remaining scales (see Table 4.2) are scored with Likert-type scales by computing a mean of the constituent item scores. Note that most scales appear on both the IMT and in-unit versions of the scales, though the IMT version has three unique scales and the in-unit version has an MOS Satisfaction scale.

We present the results for a selected subset of IMT ALQ scales in this report because of the large number of scales. The scales not reported on are: Normative Commitment, Army Career Intentions, Army Reenlistment Intentions, and Army Civilian Comparison. Normative Commitment was excluded because of its conceptual overlap with Attrition Cognitions. The two intentions scales and Army Civilian Comparison were excluded because Soldiers' attitudes at this early stage of their careers are less predictive of their actual retention behavior than attitudes captured at a later point more proximal to their behavior. Interpretation of the Army Civilian Comparison scale also varies over time as the economic climate evolves. Results for all the applicable in-unit ALQ scales are reported.

Administrative Criteria

Attrition is a broad category that encompasses involuntary and voluntary separations for a variety of reasons (e.g., underage enlistment, conduct, family concerns, drugs or alcohol, performance, physical standards or weight, mental disorder, or violations of the Uniform Code of Military Justice [UCMJ]). The reason for separation was determined by the Soldier's Separation Program Designator (SPD) code. Soldiers who were classified as "attrits" for reasons outside of the Soldiers' or the Army's control were excluded in our analyses (e.g., death or serious injury incurred while performing one's duties).

Data on IMT school performance and completion were extracted from ATRRS and RITMS data files (see Chapter 2). ATRRS course information was used to determine (a) whether a Soldier graduated from or was separated from the Army during IMT and (b) whether he or she restarted during IMT. Soldiers' final AIT course grades were extracted from RITMS. Due to restricted variance in the OSUT grades (i.e., all of the grades were pass/fail), final grades from OSUT courses were excluded from the data file.

Basic Descriptive Statistics and Psychometric Properties for IMT and In-Unit Criteria

In this section we briefly summarize the basic descriptive statistics and psychometric properties of the IMT and in-unit criterion measures for Soldiers comprising the TOPS Validation Sample from Figure 2.1 (i.e., Education Tier 1 and 2, non-prior service, AFQT Category IV or above Soldiers with valid TAPAS scores and matching criterion data). Basic descriptive statistics and psychometric

Table 4.2. ALQ Likert-Type Scales

Scale Name	Description	Number of Items	Example Item	Likert Scale Anchors
Affective Commitment	Measures Soldiers' emotional attachment to the Army.	7	I feel like I am part of the Army 'family.'	1 (strongly disagree) to 5 (strongly agree)
Normative Commitment ^a	Measures Soldiers' feelings of obligation toward staying in the Army until the end of their current term of service.	5	I would feel guilty if I left the Army before the end of my current term of service.	1 (strongly disagree) to 5 (strongly agree)
Career Intentions	Measures Soldiers' intentions to reenlist and to make the Army a career.	3	How likely is it that you will make the Army a career?	Varies by item: 1 (strongly disagree) to 5 (strongly agree); 1 (not at all confident) to 5 (extremely confident); 1 (extremely unlikely) to 5 (extremely likely)
Reenlistment Intentions	Measures Soldiers' intention to reenlist in the Army.	4	How likely is it that you will leave the Army after completing your current term of service?	1 (strongly disagree) to 5 (strongly agree)
Attrition Cognitions	Measures the degree to which Soldiers think about attriting before the end of their first term.	4	How likely is it that you will complete your current term of service?	Varies by item: 1 (strongly disagree) to 5 (strongly agree); 1 (never) to 5 (very often)
MOS Satisfaction ^b	Measures Soldiers' satisfaction with their MOS.	6	Given my skills and abilities, I think I am in the right MOS.	1 (strongly disagree) to 5 (strongly agree)
Army Life Adjustment ^a	Measures Soldiers' transition from civilian to Army life.	9	Looking back, I was not prepared for the challenges of training in the Army.	1 (strongly disagree) to 5 (strongly agree)
Army Civilian Comparison ^a	Measures Soldiers' impressions of how Army life compares to civilian life.	6	Indicate how you believe conditions in the Army compare to conditions in a civilian job with regards to pay and other factors (e.g., advancement opportunities, job security).	1 (much better in the Army) to 5 (much better in civilian life)
MOS Fit	Measures Soldiers' perceived fit with their MOS.	9	My MOS provides the right amount of challenge for me.	1 (strongly disagree) to 5 (strongly agree)
Army Fit ^c	Measures Soldiers' perceived fit with the Army.	8	The Army is a good match for me.	1 (strongly disagree) to 5 (strongly agree)

^aAppears only on the IMT ALQ.^bAppears only on the in-unit ALQ.^cScale has 6 items on the in-unit ALQ.

properties of the IMT criterion measures for the Full IMT Sample ($n = 36,401$) are reported in Appendix B. See Appendix C for corresponding statistics of the in-unit criterion measures based on the Full In-Unit Sample ($n = 1,861$). Overall, the basic descriptive statistics and reliability estimates observed in the validation samples were generally comparable to those for the full samples.

Job Knowledge Tests (JKTs)

A single, overall raw score was computed for each JKT by summing the total number of points Soldiers earned across the JKT items. All of the multiple-choice items were worth one point. Depending on the format of the non-traditional items (e.g., multiple response), they were worth one or more points. To facilitate comparisons across MOS, we computed a percent correct score based on the maximum number of points that could be obtained on each MOS test. For the criterion-related validity analyses, we converted the total raw score to a standardized score (or z -score) by standardizing the scores *within* each MOS.

JKT scores were flagged as not useable if the Soldier (a) omitted more than 10% of the assessment items, (b) took fewer than 5 minutes to complete the entire assessment, or (c) selected an implausible response to one of the careless responding items.¹⁰

Table 4.3 summarizes the percent correct scores, internal consistency reliability estimates, and correlations with AFQT scores for the MOS-specific and the WTBD JKTs by setting and education tier. Note, the psychometric properties for the in-unit MOS-specific JKTs are not reported because of insufficient sample size.

Performance Rating Scales (PRS)

For the MOS-specific PRS, a composite score was computed by averaging ratings across the individual scales. Ratings on a selected subset of Army-wide PRS were similarly combined into several unit-weighted ratings composites (as depicted in Figure 4.1), consistent with performance models used in previous Army research (Ingerick, Diaz, & Putka, 2009). PRS that were not formed into composites (i.e., Physical Fitness and Bearing, Commitment and Adjustment to the Army, and Overall Performance) were analyzed separately. The ratings composites were constructed to be as comparable as possible across the IMT and in-unit PRS in order to facilitate comparisons between the two.

A Soldier's PRS ratings were dropped if the rater provided a familiarity rating of 1 ("I have had little opportunity to observe this Soldier"). PRS data also were flagged as unusable if the rater omitted more than 10% of the assessment items or indicated that he or she had "not observed" the Soldier on more than 50% of the dimensions. Data also were removed if a rater engaged in "flat responding"—that is, ratings were removed from the data file if a rater rated 10 or more Soldiers on a particular scale and 90% or more of those rating profiles were exactly the same. Approximately 20% of Soldiers with IMT PRS were rated by more than one cadre member. None

¹⁰ The 5-minute criterion was established during the first in-unit phase of the Army Class project, which employs highly similar assessments administered via the same platform. See Knapp, Owens, et al. (2011) for details.

of the Soldiers with in-unit PRS were rated by more than one cadre member. Accordingly, interrater reliability estimates could not be estimated for the in-unit PRS.

Table 4.3. Descriptive Statistics and Reliability Estimates for IMT and In-Unit WTBD and MOS-Specific Job Knowledge Tests (JKTs) by Education Tier in the Validation Sample

Domain/Setting/MOS	<i>n</i>	<i>M</i>	<i>SD</i>	<i>Min</i>	<i>Max</i>	<i>r</i> _{WTBD}	<i>r</i> _{AFQT}	α
Tier 1								
<i>MOS-Specific (IMT)</i>								
11B/C/X + 18X	2,927	61.94	10.23	25.6	86.0	.56	.43	.77
19K	78	60.49	9.86	30.0	78.0	.52	.54	.66
31B	1,365	69.55	8.17	40.8	91.3	.48	.46	.75
68W	1,510	73.74	11.18	31.5	92.4	.53	.29	.75
88M	1,085	63.19	10.75	31.9	88.9	.56	.46	.89
91B	204	57.29	14.11	27.8	90.7	.59	.27	.91
<i>All MOS Combined</i>	7,169	65.92	11.46	25.6	92.4	.57	.46	--
<i>WTBD (Army-Wide)</i>								
IMT	8,459	64.97	13.24	6.5	96.8	--	.44	.67
In-Unit	368	70.74	12.42	23.1	96.2	--	.48	.56
Tier 2								
<i>MOS-Specific (IMT)</i>								
11B/C/X + 18X	91	60.72	10.11	30.2	76.7	.56	.33	.74
<i>All MOS Combined</i>	222	64.09	12.06	27.8	87.4	.60	.37	--
<i>WTBD (Army-Wide)</i>								
IMT	270	63.81	13.13	16.1	93.5	--	.35	.67
In-Unit	--	--	--	--	--	--	--	--
Tier 1 + 2 (Combined)								
<i>MOS-Specific (IMT)</i>								
11B/C/X + 18X	3,018	61.91	10.23	25.6	86.0	.56	.43	.77
19K	83	60.27	9.83	30.0	78.0	.52	.53	.66
31B	1,404	69.59	8.17	40.8	91.3	.48	.46	.75
68W	1,547	73.67	11.24	31.5	92.4	.53	.29	.76
88M	1,126	63.13	10.78	31.9	88.9	.57	.46	.89
91B	213	57.27	14.29	27.8	90.7	.59	.25	.91
<i>All MOS Combined</i>	7,391	65.86	11.48	25.6	92.4	.57	.46	--
<i>WTBD (Army-Wide)</i>								
IMT	8,729	64.93	13.24	6.5	96.8	--	.44	.67
In-Unit	377	70.53	12.43	23.1	96.2	--	.49	.56

Note. Means, *SDs*, *Min*, and *Max* reflect percent correct. α = coefficient alpha. WTBD = Warrior Tasks and Battle Drills. r_{WTBD} = correlation with WTBD JKT scores. r_{AFQT} = correlation with AFQT scores. All correlations are statistically significant ($p < .05$, one-tailed). Statistics based on fewer than 50 cases (which include all Tier 2) are not separately reported.

Table 4.4 summarizes the basic descriptive statistics and reliability estimates (internal consistency and interrater reliability) for the IMT PRS by domain and education tier. Mean ratings were consistently above the mid-point; a common finding in research involving performance ratings. The IMT PRS were also highly intercorrelated (see Appendix B). Table 4.5 reports the basic descriptive statistics and reliability estimates (internal consistency only) for the in-unit Army-wide PRS by domain and education tier. Overall, ratings on the in-unit PRS exhibited greater variability, on average, than the IMT PRS.

Table 4.4. Descriptive Statistics and Reliability Estimates for IMT Army-Wide and MOS-Specific Performance Rating Scales (PRS) by Education Tier in the Validation Sample

Domain/Setting/PRS	<i>n</i>	<i>M</i>	<i>SD</i>	<i>Min</i>	<i>Max</i>	<i>α</i>	IRR
Tier 1							
<i>Army-Wide</i>							
Can Do ^a	2,825	5.10	1.05	1.00	7.00	.89	--
Commitment & Adjustment	2,841	5.22	1.16	1.00	7.00	--	--
Effort & Personal Discipline ^a	2,846	5.13	1.11	1.00	7.00	.84	--
Physical Fitness & Bearing	2,832	5.06	1.15	1.00	7.00	--	--
Working with Others ^a	2,835	5.03	1.13	1.00	7.00	.84	--
Overall Performance	2,828	3.55	.82	1.00	5.00	--	--
Army-Wide ^a	2,846	5.10	1.00	1.00	7.00	.95	--
<i>MOS-Specific</i>							
11B/C/X + 18X	934	5.04	.91	1.75	7.00	.95	--
19K	55	5.26	.62	3.29	6.86	.86	--
31B	570	5.13	.92	2.13	7.00	.89	--
68W	570	4.69	.88	1.00	7.00	.96	--
88M	117	4.83	.94	2.00	7.00	.95	--
<i>All MOS Combined</i>	2,277	4.96	.93	1.00	7.00	--	--
Tier 2							
<i>Army-Wide</i>							
Can Do ^a	89	5.20	1.02	3.50	7.00	.91	--
Commitment & Adjustment	89	5.15	1.14	2.00	7.00	--	--
Effort & Personal Discipline ^a	89	4.91	1.13	2.00	7.00	.78	--
Physical Fitness & Bearing	89	4.75	1.16	2.00	7.00	--	--
Working with Others ^a	89	4.88	1.20	1.25	7.00	.82	--
Overall Performance	88	3.46	.79	1.00	5.00	--	--
Army-Wide ^a	89	4.99	.98	2.56	7.00	.93	--
<i>MOS-Specific</i>							
<i>All MOS Combined</i>	64	4.77	0.94	2.00	7.00	--	--
Tier 1 + Tier 2 (Combined)							
<i>Army-Wide</i>							
Can Do ^a	2,914	5.11	1.05	1.00	7.00	.89	.07
Commitment & Adjustment	2,930	5.22	1.16	1.00	7.00	--	.16
Effort & Personal Discipline ^a	2,935	5.12	1.11	1.00	7.00	.83	.24
Physical Fitness & Bearing	2,921	5.05	1.16	1.00	7.00	--	.22
Working with Others ^a	2,924	5.03	1.13	1.00	7.00	.84	.13
Overall Performance	2,916	3.55	.82	1.00	5.00	--	.34
Army-Wide ^a	2,935	5.09	1.00	1.00	7.00	.95	.18

Note. Statistics based on fewer than 50 cases are not reported. Ratings on PRS range from 1 and 7, except for the Overall Performance PRS, which ranges from 1 to 5. PRS ratings from supervisors with a familiarity rating of 1 (“I have had little opportunity to observe this Soldier”) were excluded from analyses. α = coefficient alpha. IRR = Interrater reliability, estimated using G(q,k) (Putka, Le, McCloy, & Diaz, 2008). IRR estimates were not estimated if 30 or fewer Soldiers were rated by more than one supervisor.

^a Ratings composite comprised of two or more Army-wide PRS (see Figure 4.1).

Table 4.4. (Continued)

Domain/Setting/PRS	<i>n</i>	<i>M</i>	<i>SD</i>	<i>Min</i>	<i>Max</i>	<i>α</i>	IRR
<i>MOS-Specific</i>							
11B/C/X + 18X	961	5.04	.91	1.75	7.00	.95	.22
19K	59	5.28	.61	3.29	6.86	.86	.52
31B	581	5.12	.92	2.13	7.00	.95	.11
68W	585	4.68	.89	1.00	7.00	.96	.00
88M	123	4.82	.95	2.00	7.00	.95	.00
<i>All MOS Combined</i>	2,341	4.95	.93	1.00	7.00	--	--

Note. Statistics based on fewer than 50 cases are not reported. Ratings on PRS range from 1 and 7, except for the Overall Performance PRS, which ranges from 1 to 5. PRS ratings from supervisors with a familiarity rating of 1 (“I have had little opportunity to observe this Soldier”) were excluded from analyses. α = coefficient alpha. IRR = Interrater reliability, estimated using G(q,k) (Putka, Le, McCloy, & Diaz, 2008). IRR estimates were not estimated if 30 or fewer Soldiers were rated by more than one supervisor.

^a Ratings composite comprised of two or more Army-wide PRS (see Figure 4.1).

As reported in Table 4.4 the interrater reliability estimates were quite low. The estimates range from .07 to .34 for the Army-wide scales in the full sample. The low estimates on the MOS-specific PRS composites are particularly problematic given these are composites based on multiple PRS. We attribute these low coefficients to several interrelated issues. First, the number of ratees per rater is high, averaging about 14 for the Full IMT Sample. As a result, raters likely became fatigued during the rating task. Second, within-rater variance was generally limited, perhaps reflecting raters’ lack of familiarity with individual Soldiers. Third, these data collections were not proctored, unlike prior research (e.g., Knapp & Heffner, 2009, 2010). Finally, the number of raters per ratee was small, averaging less than two, which limits the generalizability of single-rater reliability estimates. Although not all of these potential issues with the PRS can be addressed within the practical constraints of the research (e.g., collecting ratings in an unproctored setting), the interrater reliability may be improved by PRS format changes which were introduced in fall 2011. Results using the revised measures will appear in the next evaluation report.

Army Life Questionnaire (ALQ)

In most cases, ALQ subscale scores were computed by taking the mean of all responses associated with each scale, properly accounting for reverse coded items. The number of Training Failures, Training Achievements, Accelerated Development, and Disciplinary Actions scales were computed by summing the total number of “yes” responses. Awards Earned was computed based on the number of promotion points a Soldier earns for the awards he or she has achieved, consistent with current Army Enlisted promotion policy (Army Regulation 600-8-19). ALQ data were flagged as unusable if the Soldier (a) omitted more than 10% of the assessment items, (b) took fewer than 5 minutes to complete the entire assessment, or (c) chose an implausible response to the careless responding item.

Table 4.5. Descriptive Statistics and Reliability Estimates for In-Unit Army-Wide Performance Rating Scales (PRS) in the Validation Sample

Domain/PRS	<i>n</i>	<i>M</i>	<i>SD</i>	<i>Min</i>	<i>Max</i>	<i>α</i>
Tier 1						
<i>Army-Wide</i>						
Can Do ^a	211	4.81	1.23	1.25	7.00	.89
Effort & Personal Discipline ^a	211	5.12	1.37	1.50	7.00	.80
Physical Fitness & Bearing	210	5.21	1.58	1.00	7.00	--
Self-Management ^a	211	5.24	1.14	1.67	7.00	.78
Working with Others ^a	211	5.18	1.23	2.00	7.00	.67
Overall Leadership Potential	208	4.85	1.72	1.00	7.00	--
Army-Wide ^a	211	5.07	1.13	1.92	7.00	.95
Tier 1 + Tier 2 (Combined) ^b						
<i>Army-Wide</i>						
Can Do ^a	218	4.81	1.22	1.25	7.00	.88
Effort & Personal Discipline ^a	218	5.10	1.38	1.50	7.00	.80
Physical Fitness & Bearing	217	5.22	1.56	1.00	7.00	--
Self-Management ^a	218	5.23	1.15	1.67	7.00	.78
Working with Others ^a	218	5.17	1.23	2.00	7.00	.66
Overall Leadership Potential	215	4.83	1.73	1.00	7.00	--
Army-Wide ^a	218	5.06	1.13	1.92	7.00	.95

Note. Ratings on PRS range from 1 and 7. PRS ratings from supervisors with a familiarity rating of 1 (“I have had little opportunity to observe this Soldier”) were excluded from analyses. α = coefficient alpha. IRR = Interrater reliability, estimated using G(q,k) (Putka, Le, McCloy, & Diaz, 2008). IRR estimates were not estimated if 30 or fewer Soldiers were rated by more than one supervisor.

^a Ratings composite comprised of two or more Army-wide PRS (see Figure 4.4).

^b Tier 2 not reported separately since the sample size is less than 50 for all PRS scores.

Table 4.6 summarizes the basic descriptive statistics and internal consistency reliability estimates for the IMT ALQ scales by domain and education tier. Table 4.7 similarly reports the basic descriptive statistics and internal consistency reliability estimates for the in-unit ALQ scales by domain and education tier. See Table 4.2 for scale anchors, number of items per scale, and sample items for the ALQ.

Administrative Criteria

Administrative criterion data included both measures aimed at capturing separation-related information as well as performance information. Separation (attrition) data were obtained for Regular Army Soldiers only due to the unavailability at the time of reliable data sources for Soldiers in the reserve components. Attrition encompasses involuntary and voluntary separations before the end of the enlistment term for a variety of reasons (e.g., underage enlistment, conduct, family concerns, sexual orientation, drugs or alcohol, performance, physical standards or weight, mental disorder, or violations of the Uniformed Code of Military Justice). Soldiers who separated were coded as 1 (attrit). Soldiers who attrited for reasons that were categorized as being beyond their or the Army's control were excluded in our analyses (e.g., death or serious injury). The current analyses cover attrition through 18 months of service.

Table 4.6. Descriptive Statistics and Reliability Estimates for the IMT Army Life Questionnaire (ALQ) by Education Tier in the Validation Sample

Domain/Setting/Scale	<i>n</i>	<i>M</i>	<i>SD</i>	<i>Min</i>	<i>Max</i>	<i>α</i>
Tier 1						
<i>Retention</i>						
Affective Commitment	8,862	3.87	.67	1.00	5.00	.86
Army Life Adjustment	8,862	4.10	.66	1.00	5.00	.87
Army Fit	8,862	4.07	.60	1.00	5.00	.86
Attrition Cognitions	8,862	1.52	.60	1.00	5.00	.77
MOS Fit	8,862	3.78	.84	1.00	5.00	.92
<i>Achievement/Performance</i>						
Disciplinary Incidents	7,250	.25	.59	.00	6.00	--
Last APFT Score	8,772	251.82	30.89	118.00	300.00	--
Training Achievement	8,852	.42	.62	.00	2.00	--
Training Restarts	8,861	.36	.60	.00	4.00	--
Tier 2						
<i>Retention</i>						
Affective Commitment	281	3.96	.67	1.00	5.00	.86
Army Life Adjustment	281	4.07	.64	2.11	5.00	.86
Army Fit	281	4.11	.60	2.00	5.00	.85
Attrition Cognitions	281	1.50	.57	1.00	3.50	.71
MOS Fit	281	3.73	.83	1.11	5.00	.91
<i>Achievement/Performance</i>						
Disciplinary Incidents	236	.27	.60	.00	3.00	--
Last APFT Score	276	245.84	32.39	140.00	300.00	--
Training Achievement	281	.38	.57	.00	2.00	--
Training Restarts	281	.35	.59	.00	3.00	--
Tier 1 + Tier 2 (Combined)						
<i>Retention</i>						
Affective Commitment	9,143	3.87	.67	1.00	5.00	.86
Army Life Adjustment	9,143	4.10	.66	1.00	5.00	.87
Army Fit	9,143	4.07	.60	1.00	5.00	.86
Attrition Cognitions	9,143	1.51	.60	1.00	5.00	.77
MOS Fit	9,143	3.78	.84	1.00	5.00	.92
<i>Achievement/Performance</i>						
Disciplinary Incidents	7,486	.25	.59	.00	6.00	--
Last APFT Score	9,048	251.64	30.95	118.00	300.00	--
Training Achievement	9,133	.42	.62	.00	2.00	--
Training Restarts	9,142	.36	.60	.00	4.00	--

Note. α = coefficient alpha.

Table 4.7. Descriptive Statistics and Reliability Estimates for the In-Unit Army Life Questionnaire (ALQ) by Education Tier in the Validation Sample

Domain/Setting/Scale	<i>n</i>	<i>M</i>	<i>SD</i>	<i>Min</i>	<i>Max</i>	<i>α</i>
Tier 1						
<i>Retention</i>						
Affective Commitment	369	3.68	.75	1.00	5.00	.87
Army Career Intentions	369	2.69	1.11	1.00	5.00	.92
Army Fit	369	4.02	.65	1.83	5.00	.80
Attrition Cognitions	369	1.57	.65	1.00	5.00	.75
MOS Fit	368	3.31	.93	1.00	5.00	.93
MOS Satisfaction	369	3.56	.84	1.00	5.00	.91
Reenlistment Intentions	369	3.02	1.09	1.00	5.00	.78
<i>Achievement/Performance</i>						
Accelerated Development	367	.23	.51	0.00	3.00	--
Awards Earned (Weighted)	370	2.59	6.56	0.00	35.00	--
Disciplinary Incidents	369	.18	.49	0.00	7.00	--
Last APFT Score	357	244.74	29.45	130.00	300.00	--
Tier 1 + Tier 2 (Combined)						
<i>Retention</i>						
Affective Commitment	378	3.69	.76	1.00	5.00	.87
Army Career Intentions	378	2.71	1.13	1.00	5.00	.92
Army Fit	378	4.02	.65	1.83	5.00	.80
Attrition Cognitions	378	1.57	.66	1.00	5.00	.76
MOS Fit	377	3.31	.93	1.00	5.00	.93
MOS Satisfaction	378	3.57	.85	1.00	5.00	.91
Reenlistment Intentions	378	3.03	1.10	1.00	5.00	.78
<i>Achievement/Performance</i>						
Accelerated Development	376	.23	.51	0.00	3.00	--
Awards Earned (Weighted) ^b	379	2.57	6.52	0.00	35.00	--
Disciplinary Incidents	378	.20	.62	0.00	7.00	--
Last APFT Score	366	244.94	29.33	130.00	300.00	--

Note. α = coefficient alpha.

^aTier 2 estimates not reported since all samples sizes are less than 50.

^bAwards earned are weighted by the number of promotion points associated with each award according to current Army Enlisted promotion policy.

Administrative criterion data also included information about Soldiers' performance during IMT and was available for all Soldiers, regardless of component. For the variable, Restarted from IMT, Soldiers who graduated IMT but who had to restart at least once were coded as 1 (restart). However, Soldiers who had not had an opportunity to fully complete their IMT at the time the data were collected were excluded from our analyses. Final AIT Grades represent the final cumulative grade administratively recorded on the Soldier upon graduation from AIT. A standardized version of Final AIT Grades was computed for those MOS graduating 15 or more Soldiers. Table 4.8 summarizes the basic descriptive statistics for the administrative criteria by education tier.

Table 4.8. Base Rates or Descriptive Statistics for Administrative Attrition and IMT Criteria by Education Tier in the Validation Sample

Domain/Measure	Tier 1			Tier 2			Tier 1 + Tier 2 (Combined)		
	N^a	N_{Attrit}	$\%Attrit$	N^a	N_{Attrit}	$\%Attrit$	N^a	N_{Attrit}	$\%Attrit$
<i>Attrition</i>									
3-Month Cumulative	39,947	2,429	6.1	782	69	8.8	40,729	2,498	6.1
6-Month Cumulative	32,715	3,056	9.3	507	64	12.6	33,222	3,120	9.4
9-Month Cumulative	24,471	2,654	10.8	291	44	15.1	24,672	2,698	10.9
12-Month Cumulative	18,775	2,246	12.0	223	38	17.0	18,998	2,284	12.0
15-Month Cumulative	11,184	1,465	13.1	162	35	21.6	11,346	1,500	13.2
18-Month Cumulative	6,104	871	14.3	121	27	22.3	6,225	898	14.4
<i>Restarted Initial Military Training (IMT)</i>									
	N^b	$N_{Restarted}$	$\%Restarted$	N^b	$N_{Restarted}$	$\%Restarted$	N^b	$N_{Restarted}$	$\%Restarted$
At Least Once During IMT	32,812	4,117	12.5	977	143	14.6	33,789	4,260	12.6
For Academic or Other Pejorative Reasons	32,249	3,549	11.0	962	127	13.2	33,211	3,676	11.1
For Academic Reasons	31,682	2,986	9.4	927	93	10.0	32,609	3,079	9.4
<i>Final AIT School Grades</i>									
	N^c	M	SD	N^c	M	SD	N^c	M	SD
Overall Average (Unstandardized)	14,818	90.90	9.39	387	90.15	9.12	15,205	90.88	9.39
Overall Average (Standardized within MOS)	14,707	0.05	0.81	384	-0.12	0.84	15,091	0.05	0.81

^a N = number of Regular Army Soldiers with attrition data at the time data were extracted. N_{Attrit} = number of Soldiers who attrited at 3, 6, 9, 12, 15, and 18 months of service.

$\%Attrit$ = percentage of Soldiers who attrited through 3, 6, 9, 12, 15, and 18 months of service $[(N_{Attrit}/N) \times 100]$.

^b N = number of Soldiers with IMT data at the time data were extracted. $N_{Restarted}$ = number of Soldiers who restarted at least once during IMT. $\%Restarted$ = percentage of Soldiers who restarted at least once during IMT $[(N_{Restarted}/N) \times 100]$.

^c N = number of Soldiers with AIT school grade data at time data were extracted. Standardized school grades were not computed for MOS with insufficient sample size ($n < 15$).

Summary

Three types of criterion measures were adapted from previous Army research to validate the TAPAS: (a) JKTs, (b) PRS, and (c) the ALQ. The JKTs are completed by Soldiers in eight target MOS and measure MOS-specific and WTBD declarative and procedural knowledge. The PRS are completed by trainer cadre or supervisors and measure MOS-specific competence (IMT Soldiers only) and Army-wide constructs such as effort and leadership. Finally, the ALQ asks Soldiers to complete self-report verifiable performance items (e.g., their APFT scores) and attitudinal items (e.g., Adjustment to Army life). In general, the criterion measures exhibited acceptable and theoretically consistent psychometric properties. The exception to this was the Army-wide and MOS-specific PRS, which exhibited very low interrater reliability coefficients. Until improvements to the PRS are implemented (as discussed in Chapter 6) and reflected in the analysis data files, results concerning these scales should be interpreted with caution. Additional criterion data, such as attrition, training restarts, and AIT course grades were gathered from administrative records.

CHAPTER 5: EVIDENCE FOR THE PREDICTIVE VALIDITY OF THE TAPAS

Joseph P. Caramagno and Michael J. Ingerick (HumRRO)

This chapter summarizes the TAPAS' potential to predict Soldiers' performance and attrition during training and their first unit of assignment. We begin with a brief description of our analytic approach. Next, we summarize the main findings from (a) incremental validity analyses involving the existing TOPS composites (Can-Do, Will-Do) and constituent TAPAS scales and (b) bivariate correlations between the TAPAS scales and selected performance and retention criteria.¹¹

Analysis Approach

To evaluate the TAPAS' potential to enhance new Soldier selection, we examined its incremental validity over the AFQT in predicting first-term outcomes important to the Army. Consistent with the Army's personnel goals, we selected performance and retention-related outcomes that, as a group, provided representative coverage of the criterion space. Specific outcome measures were selected based on extant models of first-term Soldier performance and retention (Campbell, Hanson, & Oppler, 2001; Campbell, McHenry, & Wise, 1990; Knapp & Tremble, 2007; Strickland, 2005).¹²

Our approach to analyzing the TAPAS' incremental predictive validity was generally consistent with previous evaluations of the measure or similar experimental non-cognitive predictors (Ingerick, Diaz, & Putka, 2009; Knapp & Heffner, 2009; 2010; Trippe, Caramagno, Allen, & Ingerick, 2011). In brief, this approach involved testing a series of hierarchical regression models, where scores on each criterion measure were regressed onto Soldiers' AFQT scores in the first step, followed by scores on the existing TOPS composites and TAPAS scales in the second step. The resulting increment in the multiple correlation (ΔR) associated with the TAPAS served as the index of incremental validity.

For the continuously scaled criteria, the models were estimated using Ordinary Least Squares (OLS) regression. Alternatively, logistic regression was used for the dichotomous criteria (e.g., 6-, 12-, and 18-month attrition). For the logistic regression models, we estimated point-biserial correlations (r_{pb}) in place of the traditional pseudo- R estimates. These point-biserial correlations reflected the correlation between a Soldiers' predicted probability of exhibiting a selected behavior and his or her actual behavior (e.g., attriting). We estimated these correlations because of the well-known conceptual and statistical issues associated with traditional pseudo- R estimates.

¹¹ Alternative TOPS composites have been explored (Knapp & Heffner, 2012). However, at present, these alternative composites are still under development. Accordingly, the current analyses focus on the existing TOPS composites being used operationally for screening.

¹² Readers are reminded that the interrater reliability estimates for the PRS were generally low. Therefore, the predictive validity evidence associated with the PRS should be interpreted with caution. The PRS have been revised to mitigate this issue.

A series of five regression models were estimated for each criterion measure: (a) a model consisting of all TAPAS scales ($k = 15$), (b) the existing TOPS Can-Do composite ($k = 1$), (c) the existing TOPS Will-Do composite ($k = 1$), (d) a model consisting of the TAPAS scales constituting the Can Do composite ($k = 5$), and (e) a model consisting of the TAPAS scales constituting the Will Do composite ($k = 5$).¹³ Table 5.1 provides a summary of each of the regression models. All models were estimated separately by education tier, in addition to Tiers 1 and 2 combined.

Table 5.1. Summary of the Regression Models

Model Name	# of Predictors (k)	Description
All TAPAS Scales	15	Model consists of an optimally weighted composite of scores on all 15 TAPAS dimensions (or facets).
TOPS Can-Do Composite	1	Model consists of scores on the existing TOPS Can-Do composite.
TOPS Will-Do Composite	1	Model consists of scores on the existing TOPS Will-Do composite.
TAPAS Can-Do Scales	5	Model consists of an optimally weighted composite of scores on the five TAPAS dimensions (or facets) that constitute the existing Can-Do Composite.
TAPAS Will-Do Scales	5	Model consists of an optimally weighted composite of scores on the five TAPAS dimensions (or facets) that constitute the existing Will-Do Composite.

To enable comparisons across the different models and education tier levels, we adjusted the observed R and ΔR for shrinkage. Comparing the aforementioned models directly, minus the shrinkage adjustments, would have been challenging and could have led to incorrect conclusions. The reason for this is because the observed R is inflated (upwards) as more predictors are entered in the model or the sample size is small, regardless of the “true” relationship between the predictors and the criterion. This over inflation can be particularly problematic when both conditions are present – a large number of predictors and small sample size. Accordingly, adjusting the estimates for shrinkage enables a direct comparison across models for the same criterion or by education tier. We adjusted the observed R and ΔR estimates using Burket’s (1964) formula for shrinkage (cf. Formula 8; Schmitt & Ployhart, 1999). This adjustment was implemented as follows:

1. Using the observed (unadjusted) correlations among the TAPAS, AFQT, and the selected criterion previously estimated, correlations between the predictors and the performance-related criteria (e.g., JKT) were adjusted for sample size and number of predictors using Burket’s (1964) formula for shrinkage

$$\rho_c = (NR^2 - k)/[R(N - k)] \quad (1)$$

where k equals the number of elements in the model.

¹³ Analyses were limited to Soldiers with valid TAPAS 15D-Static or 15D-CAT score data (June 2009-July 2011).

2. The shrinkage-adjusted incremental validity estimates for the experimental predictors were calculated by subtracting the adjusted R associated with an AFQT-only model from the adjusted R obtained from the full model (i.e., AFQT + Experimental Predictor model).

In addition to the incremental validity analyses, we examined the predictive validity of the individual TAPAS scales based on the bivariate correlations between scores on the TAPAS scales and the targeted criterion measures.

Findings

Tables 5.2 to 5.12 summarize the main findings from the incremental validity analyses of the existing TOPS composites and constituent TAPAS scales. The results are organized by (a) setting (i.e., IMT, in-unit), (b) education tier (i.e., Tier 1, Tier 2, Tier 1 + 2 (combined)), and (c) model (i.e., all TAPAS scales, TOPS Can-Do or Will-Do composites, Can-Do or Will-Do scales). A few notes related to interpretation of the findings:

- The results for Tier 2 Soldiers should be interpreted with caution at this stage of the TOPS evaluation because of limited criterion data for those Soldiers. Accordingly, our discussion primarily focuses on the shrinkage adjusted results for Tier 1 and 2 Soldiers (combined).¹⁴
- The shrinkage adjusted results presented in this report should not be directly compared to previously reported results unadjusted for shrinkage.
- Most of our discussion focuses on the models involving the TOPS composites or their constituent scales because these models best evaluate how the TAPAS is currently being used operationally. Overall, the models based on an optimally weighted composite of all 15 TAPAS dimensions yielded incremental validity estimates that were comparable to or lower than the reduced models, on average, after adjusting the estimates for shrinkage.

Predicting IMT Performance and Retention

Tables 5.2 to 5.6 summarize the incremental validity results for predicting IMT performance and retention-related criteria. Consistent with previous analyses, the TOPS Can-Do composite and an optimally weighted composite of its constituent scales evidenced no notable increments over the AFQT in predicting technical IMT performance (with Adj. ΔR consistently less than .00). However, the Can-Do composite and scales did exhibit small to modest gains over the AFQT in predicting multiple non-technical and retention-related criteria. The largest gains for the TAPAS were evidenced in predicting Soldiers' attitudes towards the Army (Army Life Adjustment: Can-Do scales – Adj. ΔR = .11; Can-Do composite – Adj. ΔR = .07; Army Fit: Can-Do scales – Adj. ΔR = .10; Can Do composite – Adj. ΔR = .07) and APFT scores (Last APFT (ALQ): Can Do scales – Adj. ΔR = .07).

¹⁴ Results for Tier 1 and Tier 2 applicants combined were generally comparable to Tier 1 only results.

Table 5.2. Incremental Validity Estimates for the TAPAS and TOPS Composite Scales over the AFQT for Predicting IMT Technical Performance Criteria by Education Tier

IMT Criterion Measure / Model	Tier 1				Tier 2				Tier 1 + Tier 2 (Combined)			
	AFQT <i>R</i>	AFQT + TAPAS <i>R</i>	ΔR	Adj ΔR	AFQT <i>R</i>	AFQT + TAPAS <i>R</i>	ΔR	Adj ΔR	AFQT <i>R</i>	AFQT + TAPAS <i>R</i>	ΔR	Adj ΔR
<i>WTBD JKT</i>	<i>n</i> = 8,116				<i>n</i> = 249				<i>n</i> = 8,365			
All TAPAS Scales (15)	.44	.45	.00	.00	.34	.41	.07	.00	.44	.44	.00	.00
Can-Do (Composite) (1)	.44	.44	.00	.00	.34	.34	.00	.00	.44	.44	.00	.00
Will-Do (Composite) (1)	.44	.44	.00	.00	.34	.34	.01	.00	.44	.44	.00	.00
Can-Do (Scales) (5)	.44	.44	.00	.00	.34	.39	.06	.01	.44	.44	.00	.00
Will-Do (Scales) (5)	.44	.44	.00	.00	.34	.38	.04	.00	.44	.44	.00	.00
<i>MOS-Specific JKT</i>	<i>n</i> = 6,894				<i>n</i> = 204				<i>n</i> = 7,098			
All TAPAS Scales (15)	.37	.38	.01	.00	.28	.34	.07	.00	.37	.38	.01	.00
Can-Do (Composite) (1)	.37	.37	.00	.00	.28	.28	.00	.00	.37	.37	.00	.00
Will-Do (Composite) (1)	.37	.37	.00	.00	.28	.28	.00	.00	.37	.37	.00	.00
Can-Do (Scales) (5)	.37	.38	.00	.00	.28	.31	.03	.00	.37	.37	.00	.00
Will-Do (Scales) (5)	.37	.38	.00	.00	.28	.30	.02	.00	.37	.38	.00	.00
<i>Final AIT Grades</i>	<i>n</i> = 14,182				<i>n</i> = 361				<i>n</i> = 14,543			
All TAPAS Scales (15)	.31	.32	.01	.01	.33	.39	.06	.00	.31	.32	.01	.01
Can-Do (Composite) (1)	.31	.31	.00	.00	.33	.33	.00	.00	.31	.31	.00	.00
Will-Do (Composite) (1)	.31	.31	.00	.00	.33	.33	.00	.00	.31	.31	.00	.00
Can-Do (Scales) (5)	.31	.31	.01	.01	.33	.36	.02	.00	.31	.31	.01	.01
Will-Do (Scales) (5)	.31	.32	.01	.01	.33	.35	.02	.00	.31	.32	.01	.01

Note. The number in parentheses after the model title refers to the number of TAPAS-related scores in the model and excludes AFQT. AFQT = Armed Forces Qualification Test. AFQT + TAPAS = Multiple correlation (*R*) between the AFQT + selected TAPAS/TOPS composite scales with the targeted criterion measure. ΔR = Increment in *R* from adding the selected TAPAS/TOPS composite scales over AFQT to the regression model ([AFQT + TAPAS]—AFQT Only. Estimates in bold were statistically significant, $p < .01$ (one-tailed).

Table 5.3. Incremental Validity Estimates for the TAPAS and TOPS Composite Scales over the AFQT for Predicting IMT Effort and Disciplinary Incidence Criteria by Education Tier

IMT Criterion Measure / Model	Tier 1				Tier 2				Tier 1 + Tier 2 (Combined)			
	AFQT <i>R</i>	AFQT + TAPAS <i>R</i>	ΔR	Adj ΔR	AFQT <i>R</i>	AFQT + TAPAS <i>R</i>	ΔR	Adj ΔR	AFQT <i>R</i>	AFQT + TAPAS <i>R</i>	ΔR	Adj ΔR
<i>Effort & Personal Disc PRS</i>	<i>n</i> = 2,734				<i>n</i> = 81				<i>n</i> = 2,815			
All TAPAS Scales (15)	.08	.15	.07	.03	.11	.37	.26	.00	.07	.15	.07	.04
Can-Do (Composite) (1)	.08	.09	.01	.01	.11	.14	.04	.00	.07	.08	.01	.01
Will-Do (Composite) (1)	.08	.09	.01	.01	.11	.17	.06	.02	.07	.09	.02	.01
Can-Do (Scales) (5)	.08	.13	.05	.04	.11	.23	.12	.00	.07	.13	.05	.04
Will-Do (Scales) (5)	.08	.12	.04	.03	.11	.23	.13	.00	.07	.12	.05	.03
<i>Disciplinary Incidents (ALQ)</i>	<i>n</i> = 6,956				<i>n</i> = 215				<i>n</i> = 7,171			
All TAPAS Scales (15)	.04	.12	.08	.06	.06	.29	.24	.04	.04	.12	.08	.06
Can-Do (Composite) (1)	.04	.06	.03	.03	.06	.06	.00	.00	.04	.06	.02	.02
Will-Do (Composite) (1)	.04	.09	.05	.05	.06	.09	.03	.00	.04	.09	.05	.05
Can-Do (Scales) (5)	.04	.08	.04	.03	.06	.18	.13	.03	.04	.08	.04	.04
Will-Do (Scales) (5)	.04	.11	.07	.07	.06	.18	.12	.02	.04	.11	.07	.07
<i>Training Restarts (ALQ)</i>	<i>n</i> = 8,506				<i>n</i> = 260				<i>n</i> = 8,766			
All TAPAS Scales (15)	.06	.21	.15	.14	.03	.27	.24	.05	.06	.20	.15	.14
Can-Do (Composite) (1)	.06	.07	.01	.01	.03	.14	.12	.09	.06	.07	.01	.01
Will-Do (Composite) (1)	.06	.08	.02	.02	.03	.17	.14	.12	.06	.08	.02	.02
Can-Do (Scales) (5)	.06	.11	.05	.05	.03	.18	.16	.06	.06	.11	.05	.05
Will-Do (Scales) (5)	.06	.18	.12	.12	.03	.22	.19	.12	.06	.18	.12	.12

Note. The number in parentheses after the model title refers to the number of TAPAS-related scores in the model and excludes AFQT. AFQT = Armed Forces Qualification Test. AFQT + TAPAS = Multiple correlation (*R*) between the AFQT + selected TAPAS/TOPS composite scales with the targeted criterion measure. ΔR = Increment in *R* from adding the selected TAPAS/TOPS composite scales over AFQT to the regression model ([AFQT + TAPAS]—AFQT Only. Estimates in bold were statistically significant, $p < .01$ (one-tailed).

Table 5.4. Incremental Validity Estimates for the TAPAS and TOPS Composite Scales over the AFQT for Predicting IMT Physical Fitness Criteria by Education Tier

IMT Criterion Measure / Model	Tier 1				Tier 2				Tier 1 + Tier 2 (Combined)			
	AFQT <i>R</i>	AFQT + TAPAS <i>R</i>	ΔR	Adj ΔR	AFQT <i>R</i>	AFQT + TAPAS <i>R</i>	ΔR	Adj ΔR	AFQT <i>R</i>	AFQT + TAPAS <i>R</i>	ΔR	Adj ΔR
<i>Physical Fit & Bearing PRS</i>	<i>n</i> = 2,720				<i>n</i> = 81				<i>n</i> = 2,801			
All TAPAS Scales (15)	.07	.18	.12	.09	.12	.41	.29	.00	.06	.19	.12	.10
Can-Do (Composite) (1)	.07	.07	.00	.00	.12	.13	.00	.00	.06	.06	.00	.00
Will-Do (Composite) (1)	.07	.08	.01	.01	.12	.13	.01	.00	.06	.08	.01	.01
Can-Do (Scales) (5)	.07	.13	.06	.05	.12	.25	.13	.00	.06	.13	.06	.05
Will-Do (Scales) (5)	.07	.16	.10	.09	.12	.34	.22	.12	.06	.17	.10	.10
<i>Last APFT Score (ALQ)</i>	<i>n</i> = 8,422				<i>n</i> = 255				<i>n</i> = 8,677			
All TAPAS Scales (15)	.09	.32	.22	.22	.04	.40	.36	.26	.09	.32	.22	.22
Can-Do (Composite) (1)	.09	.09	.00	.00	.04	.10	.06	.02	.09	.09	.00	.00
Will-Do (Composite) (1)	.09	.12	.03	.03	.04	.11	.07	.04	.09	.12	.03	.03
Can-Do (Scales) (5)	.09	.15	.06	.06	.04	.25	.21	.15	.09	.16	.07	.06
Will-Do (Scales) (5)	.09	.31	.21	.21	.04	.30	.26	.23	.09	.31	.21	.21

Note. The number in parentheses after the model title refers to the number of TAPAS-related scores in the model and excludes AFQT. AFQT = Armed Forces Qualification Test. AFQT + TAPAS = Multiple correlation (*R*) between the AFQT + selected TAPAS/TOPS composite scales with the targeted criterion measure. ΔR = Increment in *R* from adding the selected TAPAS/TOPS composite scales over AFQT to the regression model ([AFQT + TAPAS]—AFQT Only. Estimates in bold were statistically significant, $p < .01$ (one-tailed).

Table 5.5. Incremental Validity Estimates for the TAPAS and TOPS Composite Scales over the AFQT for Predicting IMT Peer Support Criteria by Education Tier

IMT Criterion Measure / Model	Tier 1				Tier 2				Tier 1 + Tier 2 (Combined)			
	AFQT <i>R</i>	AFQT + TAPAS <i>R</i>	ΔR	Adj ΔR	AFQT <i>R</i>	AFQT + TAPAS <i>R</i>	ΔR	Adj ΔR	AFQT <i>R</i>	AFQT + TAPAS <i>R</i>	ΔR	Adj ΔR
<i>Working with Others PRS</i>	<i>n</i> = 2,723				<i>n</i> = 81				<i>n</i> = 2,804			
All TAPAS Scales (15)	.07	.14	.07	.04	.06	.47	.41	.06	.06	.14	.07	.04
Can-Do (Composite) (1)	.07	.08	.01	.01	.06	.10	.04	.00	.06	.08	.01	.01
Will-Do (Composite) (1)	.07	.08	.01	.01	.06	.08	.02	.00	.06	.08	.01	.01
Can-Do (Scales) (5)	.07	.12	.05	.04	.06	.30	.24	.06	.06	.11	.05	.04
Will-Do (Scales) (5)	.07	.12	.05	.03	.06	.31	.25	.08	.06	.11	.05	.04

Note. The number in parentheses after the model title refers to the number of TAPAS-related scores in the model and excludes AFQT. AFQT = Armed Forces Qualification Test. AFQT + TAPAS = Multiple correlation (*R*) between the AFQT + selected TAPAS/TOPS composite scales with the targeted criterion measure. ΔR = Increment in *R* from adding the selected TAPAS/TOPS composite scales over AFQT to the regression model ([AFQT + TAPAS]—AFQT Only. Estimates in bold were statistically significant, $p < .01$ (one-tailed).

Table 5.6. Incremental Validity Estimates for the TAPAS and TOPS Composite Scales over the AFQT for Predicting IMT Commitment and Satisfaction with the Army by Education Tier

IMT Criterion Measure / Model	Tier 1				Tier 2				Tier 1 + Tier 2 (Combined)			
	AFQT <i>R</i>	AFQT + TAPAS <i>R</i>	ΔR	Adj ΔR	AFQT <i>R</i>	AFQT + TAPAS <i>R</i>	ΔR	Adj ΔR	AFQT <i>R</i>	AFQT + TAPAS <i>R</i>	ΔR	Adj ΔR
<i>Commitment & Adjust PRS</i>		<i>n</i> = 2,729				<i>n</i> = 81				<i>n</i> = 2,810		
All TAPAS Scales (15)	.04	.12	.08	.05	.11	.49	.37	.10	.03	.12	.09	.05
Can-Do (Composite) (1)	.04	.06	.02	.02	.11	.12	.01	.00	.03	.05	.02	.02
Will-Do (Composite) (1)	.04	.06	.02	.02	.11	.13	.02	.00	.03	.06	.03	.02
Can-Do (Scales) (5)	.04	.10	.06	.05	.11	.27	.15	.00	.03	.10	.07	.06
Will-Do (Scales) (5)	.04	.09	.06	.04	.11	.26	.14	.00	.03	.09	.06	.05
<i>Army Life Adjustment (ALQ)</i>		<i>n</i> = 8,507				<i>n</i> = 260				<i>n</i> = 8,767		
All TAPAS Scales (15)	.07	.22	.15	.15	.06	.29	.23	.08	.06	.22	.16	.15
Can-Do (Composite) (1)	.07	.13	.07	.07	.06	.16	.10	.11	.06	.13	.07	.07
Will-Do (Composite) (1)	.07	.11	.04	.04	.06	.20	.14	.16	.06	.11	.04	.04
Can-Do (Scales) (5)	.07	.18	.11	.11	.06	.19	.13	.06	.06	.17	.11	.11
Will-Do (Scales) (5)	.07	.19	.12	.12	.06	.23	.17	.13	.06	.19	.13	.13
<i>Army Fit (ALQ)</i>		<i>n</i> = 8,507				<i>n</i> = 260				<i>n</i> = 8,767		
All TAPAS Scales (15)	.04	.16	.12	.11	.07	.21	.14	.00	.04	.16	.12	.11
Can-Do (Composite) (1)	.04	.11	.07	.07	.07	.12	.05	.03	.04	.11	.07	.07
Will-Do (Composite) (1)	.04	.09	.05	.05	.07	.10	.03	.01	.04	.09	.05	.05
Can-Do (Scales) (5)	.04	.14	.10	.10	.07	.15	.08	.00	.04	.14	.10	.10
Will-Do (Scales) (5)	.04	.14	.10	.10	.07	.15	.08	.00	.04	.14	.10	.10

Note. The number in parentheses after the model title refers to the number of TAPAS-related scores in the model and excludes AFQT. AFQT = Armed Forces Qualification Test. AFQT + TAPAS = Multiple correlation (*R*) between the AFQT + selected TAPAS/TOPS composite scales with the targeted criterion measure. ΔR = Increment in *R* from adding the selected TAPAS/TOPS composite scales over AFQT to the regression model ([AFQT + TAPAS]—AFQT Only. Estimates in bold were statistically significant, $p < .01$ (one-tailed).

The TOPS Will-Do composite and constituent scales similarly added little over the AFQT in predicting technical IMT performance (with Adj. ΔR consistently less than .01). The Will-Do composite evidenced its greatest gains in predicting the number of disciplinary incidents and retention-related attitudes (Adjustment to Army Life, Army Fit). However, none of the adjusted ΔR s exceeded .05. Overall, the biggest gains in predictive validity were exhibited by an optimally weighted composite of the constituent Will-Do scales. An optimally weighted composite of the Will Do scales significantly enhanced the AFQT in the prediction of physical fitness test scores (Adj. $\Delta R = .21$), attitudes towards the Army (Adjustment to Army Life, Adj. $\Delta R = .13$; Army Fit, Adj. $\Delta R = .10$), number of training restarts (ALQ Training Restarts Adj. $\Delta R = .12$), ratings of physical fitness and bearing (Physical Fitness & Bearing PRS, Adj. $\Delta R = .10$), and number of disciplinary incidents (Disciplinary Incidents, Adj. $\Delta R = .07$).

Incremental validity estimates for the TOPS composites or their constituent scales in predicting attrition at 6-, 12-, and 18-months of service were small ($\Delta R = .00 - .04$). This pattern was true for the adjusted and unadjusted estimates of the ΔR . Results were comparable for an optimally weighted composite of all 15 TAPAS scales, except for Tier 2 Soldiers. Among Tier 2 Soldiers, incremental validity estimates based on the full TAPAS were .09 or higher, after the shrinkage adjustment.

Predicting In-Unit Performance and Retention

Tables 5.8 to 5.12 summarize the incremental validity results for predicting in-unit performance- and retention-related criteria. No analyses were conducted using Tier 2 Soldiers only because of limited in-unit criterion data ($n < 10$). Consistent with the previous sections, our discussion focuses on the shrinkage adjusted results for Tier 1 and 2 Soldiers combined. The TOPS Can-Do composite exhibited meaningful gains in the prediction of in-unit Soldiers' commitment and satisfaction with the Army (Army Fit, Adj. $\Delta R = .18$; Affective Commitment, Adj. $\Delta R = .14$). Incremental validity estimates for an optimally weighted composite of the constituent Can-Do scales were comparable for the retention-related criteria (Army Fit, Adj. $\Delta R = .13$; Affective Commitment, Adj. $\Delta R = .11$) as well as Last APFT Score (Adj. $\Delta R = .12$). The TOPS Will-Do composite exhibited a small increment over the AFQT in predicting Soldiers' emotional attachments to the Army (Affective Commitment, Adj. $\Delta R = .12$). In contrast to the IMT results, however, the Will-Do composite did not significantly predict Soldiers' perceived fit with the Army. An optimally weighted composite of the Will-Do scales evidenced modest, statistically significant gains in the prediction of Last APFT Scores (Adj. $\Delta R = .28$), Army Fit (Adj. $\Delta R = .17$), and Affective Commitment (Adj. $\Delta R = .11$).

As mentioned previously, incremental validity estimates based on an optimally weighted composite of all 15 TAPAS scales were generally similar to models for the Can-Do and Will-Do constituent scales.

Table 5.7. Incremental Validity Estimates for the TAPAS and TOPS Composite Scales over the AFQT for Predicting Cumulative Attrition through 18 Months of Service by Education Tier

Criterion Measure / Model	Tier 1				Tier 2				Tier 1 + Tier 2 (Combined)			
	AFQT r_{pb}	AFQT + TAPAS r_{pb}	Δr_{pb}	Adj Δr_{pb}	AFQT r_{pb}	AFQT + TAPAS r_{pb}	Δr_{pb}	Adj Δr_{pb}	AFQT r_{pb}	AFQT + TAPAS r_{pb}	Δr_{pb}	Adj Δr_{pb}
<i>6-Month</i>	<i>n = 31,942</i>				<i>n = 479</i>				<i>n = 32,421</i>			
All TAPAS Scales (15)	.05	.10	.05	.04	.04	.23	.19	.09	.05	.10	.05	.04
Can-Do (Composite) (1)	.05	.05	.00	.00	.04	.07	.03	.01	.05	.05	.00	.00
Will-Do (Composite) (1)	.05	.06	.00	.00	.04	.05	.02	.00	.05	.06	.00	.00
Can-Do (Scales) (5)	.05	.07	.01	.01	.04	.15	.11	.07	.05	.06	.01	.01
Will-Do (Scales) (5)	.05	.09	.04	.04	.04	.13	.10	.04	.05	.09	.04	.03
<i>12-Month</i>	<i>n = 18,345</i>				<i>n = 220</i>				<i>n = 18,565</i>			
All TAPAS Scales (15)	.06	.11	.05	.04	.05	.33	.28	.12	.07	.11	.05	.04
Can-Do (Composite) (1)	.06	.07	.00	.00	.05	.06	.01	.00	.07	.07	.00	.00
Will-Do (Composite) (1)	.06	.07	.00	.00	.05	.11	.06	.02	.07	.07	.00	.00
Can-Do (Scales) (5)	.06	.08	.01	.01	.05	.14	.09	.00	.07	.08	.01	.01
Will-Do (Scales) (5)	.06	.10	.04	.04	.05	.14	.09	.00	.07	.10	.04	.03
<i>18-Month</i>	<i>n = 5,703</i>				<i>n = 118</i>				<i>n = 5,821</i>			
All TAPAS Scales (15)	.12	.15	.03	.02	.15	.56	.41	.27	.12	.15	.03	.01
Can-Do (Composite) (1)	.12	.12	.00	.00	.15	.18	.02	.00	.12	.12	.00	.00
Will-Do (Composite) (1)	.12	.12	.00	.00	.15	.30	.15	.15	.12	.12	.00	.00
Can-Do (Scales) (5)	.12	.13	.01	.00	.15	.19	.04	.00	.12	.13	.01	.00
Will-Do (Scales) (5)	.12	.14	.02	.02	.15	.40	.25	.19	.12	.14	.02	.01

Note. The number in parentheses after the model title refers to the number of TAPAS-related scores in the model and excludes AFQT. AFQT = Armed Forces Qualification Test. AFQT + TAPAS = Multiple correlation (R) between the AFQT + selected TAPAS/TOPS composite scales with the targeted criterion measure. Δr_{pb} = Increment in r_{pb} from adding the selected TAPAS/TOPS composite scales over AFQT to the regression model ([AFQT + TAPAS]—AFQT Only). Estimates are the observed *point-biserial correlation* (r_{pb}) between Soldiers' predicted probability of an event (e.g., attrition, graduating IMT without a restart) and their actual behavior. Large, positive r_{pb} values mean that the TOPS composite or scale positively predicted Soldiers' actual behavior. Estimates in bold were statistically significant, $p < .01$ (one-tailed).

Table 5.8. Incremental Validity Estimates for the TAPAS and TOPS Composite Scales over the AFQT for Predicting In-Unit Technical Performance Criteria by Education Tier

In-Unit Criterion Measure / Model	Tier 1				Tier 1 + Tier 2 (Combined)			
	AFQT <i>R</i>	AFQT + TAPAS <i>R</i>	ΔR	Adj ΔR	AFQT <i>R</i>	AFQT + TAPAS <i>R</i>	ΔR	Adj ΔR
<i>WTBD JKT</i>	<i>n</i> = 365				<i>n</i> = 373			
All TAPAS Scales (15)	.48	.55	.07	.02	.49	.56	.07	.02
Can-Do (Composite) (1)	.48	.49	.01	.00	.49	.49	.01	.00
Will-Do (Composite) (1)	.48	.50	.02	.02	.49	.51	.02	.02
Can-Do (Scales) (5)	.48	.51	.03	.01	.49	.51	.03	.01
Will-Do (Scales) (5)	.48	.51	.03	.01	.49	.52	.03	.01
<i>Can Do PRS</i>	<i>n</i> = 209				<i>n</i> = 215			
All TAPAS Scales (15)	.16	.32	.15	.00	.16	.30	.14	.00
Can-Do (Composite) (1)	.16	.17	.01	.00	.16	.17	.01	.00
Will-Do (Composite) (1)	.16	.17	.01	.00	.16	.17	.01	.00
Can-Do (Scales) (5)	.16	.21	.05	.00	.16	.20	.04	.00
Will-Do (Scales) (5)	.16	.20	.04	.00	.16	.20	.04	.00

Note. The number in parentheses after the model title refers to the number of TAPAS-related scores in the model and excludes AFQT. Tier 2 results not reported because $n < 10$. AFQT = Armed Forces Qualification Test. AFQT + TAPAS = Multiple correlation (R) between the AFQT + selected TAPAS/TOPS composite scales with the targeted criterion measure. ΔR = Increment in R from adding the selected TAPAS/TOPS composite scales over AFQT to the regression model ([AFQT + TAPAS]—AFQT Only. Estimates in bold were statistically significant, $p < .05$ (one-tailed).

Table 5.9. Incremental Validity Estimates for the TAPAS and TOPS Composite Scales over the AFQT for Predicting In-Unit Effort and Disciplinary Incidents Criteria by Education Tier

In-Unit Criterion Measure / Model	Tier 1				Tier 1 + Tier 2 (Combined)			
	AFQT <i>R</i>	AFQT + TAPAS <i>R</i>	ΔR	Adj ΔR	AFQT <i>R</i>	AFQT + TAPAS <i>R</i>	ΔR	Adj ΔR
<i>Effort & Personal Disc PRS</i>	<i>n</i> = 209				<i>n</i> = 215			
All TAPAS Scales (15)	.18	.32	.15	.00	.18	.32	.14	.00
Can-Do (Composite) (1)	.18	.18	.00	.00	.18	.18	.00	.00
Will-Do (Composite) (1)	.18	.19	.02	.00	.18	.19	.01	.00
Can-Do (Scales) (5)	.18	.23	.06	.00	.18	.23	.05	.00
Will-Do (Scales) (5)	.18	.25	.07	.00	.18	.25	.06	.00
<i>Disciplinary Incidents (ALQ)</i>	<i>n</i> = 366				<i>n</i> = 374			
All TAPAS Scales (15)	.09	.23	.14	.00	.10	.25	.15	.02
Can-Do (Composite) (1)	.09	.09	.00	.00	.10	.10	.00	.00
Will-Do (Composite) (1)	.09	.11	.03	.01	.10	.12	.02	.00
Can-Do (Scales) (5)	.09	.14	.05	.00	.10	.14	.05	.00
Will-Do (Scales) (5)	.09	.18	.09	.04	.10	.21	.11	.06
<i>Accelerated Develop(ALQ)</i>	<i>n</i> = 364				<i>n</i> = 372			
All TAPAS Scales (15)	.04	.16	.12	.00	.04	.16	.12	.00
Can-Do (Composite) (1)	.04	.04	.00	.00	.04	.05	.00	.00
Will-Do (Composite) (1)	.04	.05	.01	.00	.04	.05	.01	.00
Can-Do (Scales) (5)	.04	.10	.06	.00	.04	.11	.06	.00
Will-Do (Scales) (5)	.04	.12	.09	.00	.04	.13	.08	.00

Note. The number in parentheses after the model title refers to the number of TAPAS-related scores in the model and excludes AFQT. Tier 2 results not reported because $n < 10$. AFQT = Armed Forces Qualification Test. AFQT + TAPAS = Multiple correlation (R) between the AFQT + selected TAPAS/TOPS composite scales with the targeted criterion measure. ΔR = Increment in R from adding the selected TAPAS/TOPS composite scales over AFQT to the regression model ([AFQT + TAPAS]—AFQT Only. Estimates in bold were statistically significant, $p < .05$ (one-tailed).

Table 5.10. Incremental Validity Estimates for the TAPAS and TOPS Composite Scales over the AFQT for Predicting In-Unit Physical Fitness Criteria by Education Tier

In-Unit Criterion Measure / Model	Tier 1				Tier 1 + Tier 2 (Combined)			
	AFQT <i>R</i>	AFQT + TAPAS <i>R</i>	ΔR	Adj ΔR	AFQT <i>R</i>	AFQT + TAPAS <i>R</i>	ΔR	Adj ΔR
<i>Physical Fit & Bearing PRS</i>	<i>n</i> = 208				<i>n</i> = 214			
All TAPAS Scales (15)	.17	.39	.22	.06	.17	.38	.22	.07
Can-Do (Composite) (1)	.17	.18	.01	.00	.17	.17	.01	.00
Will-Do (Composite) (1)	.17	.18	.01	.00	.17	.18	.01	.00
Can-Do (Scales) (5)	.17	.25	.08	.00	.17	.24	.08	.00
Will-Do (Scales) (5)	.17	.28	.11	.04	.17	.28	.12	.05
<i>Last APFT (ALQ)</i>	<i>n</i> = 354				<i>n</i> = 362			
All TAPAS Scales (15)	.05	.37	.32	.26	.04	.36	.33	.25
Can-Do (Composite) (1)	.05	.05	.00	.00	.04	.04	.00	.00
Will-Do (Composite) (1)	.05	.12	.07	.07	.04	.12	.08	.07
Can-Do (Scales) (5)	.05	.20	.16	.12	.04	.20	.17	.12
Will-Do (Scales) (5)	.05	.33	.28	.28	.04	.33	.29	.28

Note. The number in parentheses after the model title refers to the number of TAPAS-related scores in the model and excludes AFQT. Tier 2 results not reported because $n < 10$. AFQT = Armed Forces Qualification Test. AFQT + TAPAS = Multiple correlation (R) between the AFQT + selected TAPAS/TOPS composite scales with the targeted criterion measure. ΔR = Increment in R from adding the selected TAPAS/TOP composite scales over AFQT to the regression model ([AFQT + TAPAS]—AFQT Only. Estimates in bold were statistically significant, $p < .05$ (one-tailed).

Table 5.11. Incremental Validity Estimates for the TAPAS and TOPS Composite Scales over the AFQT for Predicting In-Unit Peer Support and Leadership Potential Criteria by Education Tier

In-Unit Criterion Measure / Model	Tier 1				Tier 1 + Tier 2 (Combined)			
	AFQT <i>R</i>	AFQT + TAPAS <i>R</i>	ΔR	Adj ΔR	AFQT <i>R</i>	AFQT + TAPAS <i>R</i>	ΔR	Adj ΔR
<i>Work with Others PRS</i>	<i>n</i> = 209				<i>n</i> = 215			
All TAPAS Scales (15)	.17	.38	.21	.06	.16	.37	.21	.05
Can-Do (Composite) (1)	.17	.17	.00	.00	.16	.16	.00	.00
Will-Do (Composite) (1)	.17	.21	.04	.03	.16	.20	.04	.02
Can-Do (Scales) (5)	.17	.18	.01	.00	.16	.17	.01	.00
Will-Do (Scales) (5)	.17	.29	.12	.06	.16	.28	.12	.05
<i>Overall Leadership Pot PRS</i>	<i>n</i> = 206				<i>n</i> = 212			
All TAPAS Scales (15)	.17	.37	.19	.02	.17	.35	.18	.01
Can-Do (Composite) (1)	.17	.18	.00	.00	.17	.17	.00	.00
Will-Do (Composite) (1)	.17	.19	.02	.00	.17	.18	.01	.00
Can-Do (Scales) (5)	.17	.27	.10	.02	.17	.26	.09	.01
Will-Do (Scales) (5)	.17	.23	.06	.00	.17	.23	.06	.00

Note. The number in parentheses after the model title refers to the number of TAPAS-related scores in the model and excludes AFQT. Tier 2 results not reported because $n < 10$. AFQT = Armed Forces Qualification Test. AFQT + TAPAS = Multiple correlation (R) between the AFQT + selected TAPAS/TOPS composite scales with the targeted criterion measure. ΔR = Increment in R from adding the selected TAPAS/TOP composite scales over AFQT to the regression model ([AFQT + TAPAS]—AFQT Only. Estimates in bold were statistically significant, $p < .05$ (one-tailed).

Table 5.12. Incremental Validity Estimates for the TAPAS and TOPS Composite Scales over the AFQT for Predicting In-Unit Commitment and Satisfaction with the Army by Education Tier

In-Unit Criterion Measure / Model	Tier 1				Tier 1 + Tier 2 (Combined)			
	AFQT <i>R</i>	AFQT + TAPAS <i>R</i>	ΔR	Adj ΔR	AFQT <i>R</i>	AFQT + TAPAS <i>R</i>	ΔR	Adj ΔR
<i>Affective Commitment (ALQ)</i>	<i>n</i> = 366				<i>n</i> = 374			
All TAPAS Scales (15)	.01	.25	.25	.08	.01	.24	.24	.07
Can-Do (Composite) (1)	.01	.17	.17	.14	.01	.17	.16	.14
Will-Do (Composite) (1)	.01	.16	.15	.12	.01	.15	.14	.11
Can-Do (Scales) (5)	.01	.20	.19	.11	.01	.19	.19	.11
Will-Do (Scales) (5)	.01	.20	.19	.11	.01	.19	.19	.11
<i>Army Fit (ALQ)</i>	<i>n</i> = 366				<i>n</i> = 374			
All TAPAS Scales (15)	.01	.28	.27	.13	.00	.28	.28	.13
Can-Do (Composite) (1)	.01	.20	.20	.18	.00	.21	.20	.18
Will-Do (Composite) (1)	.01	.10	.09	.04	.00	.09	.09	.03
Can-Do (Scales) (5)	.01	.20	.20	.13	.00	.20	.20	.13
Will-Do (Scales) (5)	.01	.24	.23	.17	.00	.24	.23	.17
<i>Reenlistment Intent (ALQ)</i>	<i>n</i> = 366				<i>n</i> = 374			
All TAPAS Scales (15)	.02	.17	.15	.00	.03	.18	.15	.00
Can-Do (Composite) (1)	.02	.11	.09	.06	.03	.12	.10	.08
Will-Do (Composite) (1)	.02	.09	.08	.03	.03	.10	.08	.05
Can-Do (Scales) (5)	.02	.13	.11	.00	.03	.14	.12	.03
Will-Do (Scales) (5)	.02	.13	.11	.00	.03	.14	.11	.03

Note. The number in parentheses after the model title refers to the number of TAPAS-related scores in the model and excludes AFQT. Tier 2 results not reported because $n < 10$. AFQT = Armed Forces Qualification Test. AFQT + TAPAS = Multiple correlation (R) between the AFQT + selected TAPAS/TOPS composite scales with the targeted criterion measure. ΔR = Increment in R from adding the selected TAPAS/TOPS composite scales over AFQT to the regression model ([AFQT + TAPAS]—AFQT Only. Estimates in bold were statistically significant, $p < .05$ (one-tailed).

Tables 5.13 to 5.15 summarize the bivariate correlations between the AFQT, the individual TAPAS scales, and selected IMT criteria by education tier. Consistent with the incremental validity results and previous research, AFQT scores were strongly positively correlated with IMT technical performance (WTBD JKT, $r = .44$, MOS-Specific JKT, $r = .37$, Final AIT Grade, $r = .31$; $p < .01$). As expected, TAPAS scales correlated more strongly with the non-technical performance criteria, on average, than AFQT scores, although these correlations were generally smaller in magnitude (less than .20). The largest correlations were associated with Physical Conditioning scores which were positively correlated with self-reported APFT score ($r = .29$, $p < .01$) and Adjustment to Army Life ($r = .14$, $p < .01$), and negatively correlated with number of training restarts ($r = -.16$, $p < .01$). Intellectual Efficiency was positively correlated with WTBD JKT ($r = .20$, $p < .01$), MOS-specific JKT ($r = .17$, $p < .01$), Final AIT Grade ($r = .14$, $p < .01$), and Adjustment to Army Life ($r = .11$, $p < .01$). Multiple TAPAS scales correlated significantly with attrition at 6-, 12- and 18-months of service in the expected direction. However, except for Physical Conditioning (r 's = $-.06$ -. $.07$), none of the correlations were larger than $\pm .04$.

Summary

This chapter summarized results from the fourth cycle of the TOPS evaluation. Overall, the existing TOPS composites and their constituent TAPAS scales evidenced modest incremental validity over the AFQT in predicting first-term Soldier performance and retention. Incremental validity estimates (adjusted for shrinkage) were consistently .03 or less, on average, for the existing Can-Do and Will-Do composites. The constituent TAPAS scales exhibited somewhat higher incremental validity than the current composites, on average, even after adjusting for shrinkage (with Avg. Adj. ΔR ranging .04 to .06). Consistent with prior evaluation cycles, the Will-Do composite or constituent scales continued to evidence somewhat higher incremental validity, on average, than the Can-Do composite. The Will-Do composites or their constituent scales demonstrated the biggest gains over the AFQT in predicting physically-oriented performance and adjustment to and attitudes about the Army. Incremental validity gains were consistently .03 or less for predicting attrition through 18 months of service. At the scale-level, Intellectual Efficiency emerged as the single best correlate of technical performance criteria (r 's ranging from .14 to .20), while Physical Conditioning represented the single best and most consistent correlate of non-technical performance and attrition (r 's ranging from $-.09$ to .28).

Table 5.13. Summary of the Bivariate Correlations between AFQT, TAPAS, and Selected IMT Criteria for Tier 1 Soldiers

Measure/Scale	Technical Performance			Effort & Personal Discipline		Physical Fitness	Commitment & Satisfaction with the Army		Attrition		
	WTBD JKT	MOS-Specific JKT	Final AIT Grade	Disciplinary Incidents (ALQ)	Training Restarts (ALQ)	Last APFT Scores (ALQ)	Army Life Adjustment (ALQ)	Army Fit (ALQ)	6-Month	12-Month	18-Month
<i>AFQT</i>	.44	.37	.31	-.04	-.06	.09	.08	-.03	-.05	-.07	-.13
<i>Individual TAPAS Scales</i>											
Achievement	.05	.05	.08	-.07	-.07	.08	.13	.12	-.01	-.02	-.01
Adjustment	.07	.05	.01	-.01	-.05	.01	.07	.01	-.02	-.03	-.01
Attention Seeking	.04	.01	-.01	.00	-.05	.07	.06	.03	-.02	-.03	-.02
Cooperation	-.01	-.01	.01	.00	.01	-.01	.00	.01	-.01	.00	.00
Dominance	.04	.00	.03	-.05	-.10	.12	.14	.09	-.02	-.02	-.03
Even Tempered	.05	.03	.03	-.01	.03	-.06	.01	.01	-.01	-.01	-.01
Intellectual Efficiency	.20	.17	.14	-.02	-.07	.03	.11	.03	-.01	-.01	-.02
Non-Delinquency	.01	-.01	.04	-.03	.02	-.04	.00	.04	.01	.02	.02
Optimism	-.01	-.01	.01	-.03	-.04	.04	.10	.06	-.03	-.03	-.03
Order	-.06	-.07	-.03	-.01	.02	.03	-.01	.01	.02	.02	.03
Physical Conditioning	.01	-.02	.00	-.09	-.16	.29	.14	.05	-.07	-.07	-.07
Self Control	.01	.01	.02	-.01	.01	-.01	.02	.04	.00	.00	.00
Selflessness	-.02	-.03	-.01	.00	.05	.00	-.01	.05	.04	.04	.02
Sociability	-.07	-.09	-.07	.01	-.03	.04	.04	.03	.00	.01	.02
Tolerance	-.03	-.03	-.03	.01	.07	.00	.02	.04	.01	.01	.01
<i>N</i>											
AFQT	8,736	7,385	15,279	7,486	9,163	9,078	9,164	9,164	33,942	19,513	6,437
TAPAS	8,116	6,894	14,182	6,956	8,506	8,422	8,507	8,507	31,942	18,345	5,703

Note. Correlations in bold are statistically significant ($p < .01$, two-tailed).

Table 5.14. Summary of the Bivariate Correlations between AFQT, TAPAS, and Selected IMT Criteria and Attrition for Tier 2 Soldiers

Measure/Scale	Technical Performance			Effort & Personal Discipline		Physical Fitness	Commitment & Satisfaction with the Army		Attrition		
	WTBD JKT	MOS-Specific JKT	Final AIT Grade	Disciplinary Incidents (ALQ)	Training Restarts (ALQ)	Last APFT Scores (ALQ)	Army Life Adjustment (ALQ)	Army Fit (ALQ)	6-Month	12-Month	18-Month
<i>AFQT</i>	.34	.29	.33	-.05	-.01	.03	-.03	-.05	-.04	-.05	-.14
<i>Individual TAPAS Scales</i>											
Achievement	.09	.02	.05	-.10	-.16	.15	.14	.11	.03	.05	.06
Adjustment	.05	.06	.10	.00	-.05	.09	.08	.00	.05	.06	-.02
Attention Seeking	-.02	.00	.02	.07	.06	.03	-.03	.04	-.02	-.06	-.28
Cooperation	.00	.01	.10	-.06	-.10	.11	.04	.01	.03	-.08	-.14
Dominance	.00	-.06	-.01	-.01	-.04	.19	.10	.08	.01	.03	.08
Even Tempered	.17	.12	-.05	.08	-.08	-.08	.05	.00	.08	.02	.04
Intellectual Efficiency	.11	.04	.19	.04	-.09	.05	-.01	-.03	.04	-.06	-.05
Non-Delinquency	.02	.01	.06	-.02	-.01	-.01	.10	.06	-.05	.01	.02
Optimism	-.08	-.05	.00	-.05	-.06	.17	.11	.08	.06	.12	.12
Order	-.10	-.04	-.08	.13	.05	.00	.05	-.03	.00	.06	.06
Physical Conditioning	-.08	-.08	.00	-.06	-.12	.27	.16	.05	.03	.12	.24
Self Control	.06	.01	.04	.00	.00	.06	.01	.04	.13	.14	.26
Selflessness	-.02	.07	-.02	.08	-.02	.01	-.04	.01	.05	-.02	-.04
Sociability	.01	-.01	-.03	.03	-.06	.12	.06	.07	-.09	-.13	-.12
Tolerance	-.02	-.06	-.05	-.04	.02	.13	-.11	-.09	.00	-.01	-.02
<i>N</i>											
AFQT	276	226	407	242	288	283	288	288	533	235	125
TAPAS	249	204	361	215	260	255	260	260	479	220	118

Note. Correlations in bold are statistically significant ($p < .01$, two-tailed).

Table 5.15. Bivariate Correlations between AFQT, TAPAS, and Selected IMT Criteria, and Attrition for Tier 1 + Tier 2 Soldiers

Measure/Scale	Technical Performance			Effort & Personal Discipline		Physical Fitness	Commitment & Satisfaction with the Army		Attrition		
	WTBD JKT	MOS-Specific JKT	Final AIT Grade	Disciplinary Incidents (ALQ)	Training Restarts (ALQ)	Last APFT Scores (ALQ)	Army Life Adjustment (ALQ)	Army Fit (ALQ)	6-Month	12-Month	18-Month
AFQT	.44	.37	.31	-.04	-.06	.09	.07	-.03	-.05	-.07	-.13
<i>Individual TAPAS Scales</i>											
Achievement	.05	.04	.08	-.07	-.07	.08	.13	.12	-.01	-.02	-.01
Adjustment	.07	.05	.01	-.01	-.05	.01	.07	.01	-.02	-.03	-.01
Attention Seeking	.04	.01	.00	.00	-.05	.07	.06	.03	-.02	-.03	-.03
Cooperation	-.01	-.01	.01	.00	.00	.00	.00	.01	-.01	.00	.00
Dominance	.04	.00	.03	-.05	-.10	.12	.13	.09	-.02	-.02	-.03
Even Tempered	.05	.03	.03	-.01	.03	-.06	.01	.01	-.01	-.01	-.01
Intellectual Efficiency	.20	.17	.14	-.02	-.07	.02	.11	.03	-.01	-.01	-.02
Non-Delinquency	.01	-.01	.04	-.03	.03	-.04	.00	.04	.01	.02	.01
Optimism	-.01	-.01	.01	-.03	-.04	.05	.10	.06	-.02	-.03	-.02
Order	-.07	-.07	-.03	-.01	.02	.03	.00	.01	.02	.02	.03
Physical Conditioning	.01	-.02	.00	-.09	-.16	.29	.14	.05	-.06	-.07	-.07
Self Control	.02	.01	.02	-.01	.01	-.01	.02	.04	.00	.00	.00
Selflessness	-.02	-.03	-.02	.00	.05	.00	-.01	.05	.04	.04	.02
Sociability	-.07	-.09	-.07	.01	-.03	.04	.04	.03	.00	.00	.01
Tolerance	-.03	-.04	-.03	.01	.07	.00	.01	.04	.01	.01	.01
<i>N</i>											
AFQT	9,012	7,611	15,686	7,728	9,451	9,361	9,452	9,452	34,475	19,748	6,562
TAPAS	8,365	7,098	14,543	7,171	8,766	8,677	8,767	8,767	32,421	18,565	5,821

Note. Correlations in bold are statistically significant ($p < .01$, two-tailed).

CHAPTER 6: SUMMARY AND A LOOK AHEAD

Deirdre J. Knapp (HumRRO), Tonia S. Heffner, Leonard A. White, and Kate LaPort (ARI)

Summary of the TOPS IOT&E Method

In an effort to expand the basis on which applicants are evaluated for enlistment, the Army is conducting an IOT&E of the Tier One Performance Screen (TOPS). The TOPS assessments, including the Tailored Adaptive Personality Assessment System (TAPAS), the Information/Communications Technology Literacy (ICTL) test, and starting before the end of CY2012, the Work Preferences Assessment (WPA), are being administered to non-prior service applicants testing at MEPS locations.

To evaluate the TAPAS, ICTL, and WPA, the Army is collecting training criterion data on Soldiers in selected MOS as they complete their IMT. The criterion measures include job knowledge tests (JKTs), an attitudinal person-environment fit assessment (the Army Life Questionnaire; ALQ), and performance rating scales (PRS) completed by the Soldiers' cadre members. Course grades and completion rates are obtained from administrative records for all Soldiers, regardless of MOS. The plan is to construct analysis datasets and conduct cumulative validation analyses at 6-month intervals throughout the IOT&E period.

Job performance data are also being collected from Soldiers in their units to gather data on Soldiers from across all MOS who completed the TAPAS (and WPA and ICTL) at entry. These measures again include JKTs, the ALQ, and supervisor ratings. Finally, the separation status of all Soldiers who took the TAPAS at entry is being tracked throughout the course of the research.

The January 2012 data file (containing data collected through December 2011), which was the basis for analyses documented in this report, includes a total of 216,565 applicants who took the TAPAS between June 2009 and July 2011. Of these total applicants, 176,467 were in the TOPS Applicant Sample. The Applicant Sample was determined by excluding Education Tier 3, AFQT Category V, and prior service applicants from the master data file. The validation sample sizes are considerably smaller, with the IMT Validation Sample comprising 9,502 Soldiers and the Administrative Validation Sample (which includes Soldiers for whom we only have TAPAS and administrative criterion data) comprising 69,495 Soldiers.

The JKT, ALQ, and administrative criterion measures exhibited acceptable and theoretically consistent psychometric properties. The Army-wide and MOS-specific PRS, however, continued to exhibit very low interrater reliability. The PRS instruments have been revised to change both content and format in an attempt to improve their psychometric characteristics. Details of these changes will be presented when we start including data from the new measures in the analysis data files. Until improvements can be implemented, results based on supervisor ratings should be interpreted with caution.

Summary of Evaluation Results to Date

Evaluation results thus far suggest that, while the magnitude of the predictive validity estimates are not as large as those found in the experimental EEEM research (Knapp & Heffner, 2010), the TAPAS holds promise for new Soldier selection. Many of the scale-level coefficients are consistent with a theoretical understanding of the TAPAS scales, suggesting that the scales are measuring the characteristics that they are intended to measure. However, given the restricted nature of the matched criterion sample (in terms of sample characteristics) and the low reliability of the ratings data, these results should be considered preliminary.

Looking Ahead

Predictor Measures

In August 2011, three new adaptive versions of the TAPAS were introduced into the MEPS, based on a new Army-specific statement pool. Each version measures 15 dimensions. All three versions assess the same nine core dimensions, to include all of the scales in the existing TOPS “can-do” and “will-do” composites, plus six of 12 experimental dimensions (or scales). The six experimental dimensions assessed vary by version (see Table 6.1). In total, the newer versions of the TAPAS collectively measure 21 dimensions. The experimental dimensions will be evaluated for potential use in revised or new TOPS composites, once sufficient data are available.

Table 6.1. TAPAS Dimensions Assessed

	Version A	Version B	Version C
Achievement	✓	✓	✓
Adjustment	✓	✓	✓
Adventure Seeking		✓	
Attention Seeking	✓	✓	✓
Commitment to Serve		✓	
Cooperation	✓	✓	
Courage			✓
Dominance	✓	✓	✓
Even Tempered	✓	✓	✓
Intellectual Efficiency	✓	✓	✓
Non-Delinquency	✓	✓	✓
Optimism	✓	✓	✓
Order	✓	✓	
Physical Conditioning	✓	✓	✓
Responsibility			✓
Self Control	✓		✓
Selflessness	✓	✓	
Sociability	✓		✓
Situational Awareness		✓	
Team Orientation			✓
Tolerance	✓		✓

Criterion Measures

In mid-2011, the MOS-specific and WTBD JKTs (both training and in-unit versions) were reviewed and updated with the assistance of Army subject matter experts. As part of this effort, additional items were added to the WTBD JKT in an effort to increase both its reliability and content representativeness. Additional items were also added to the 31B JKTs to cover content domains that have increased in relevance since the test blueprint was originally developed. In addition to updating and improving existing measures, we developed MOS-specific measures (both IMT and in-unit) for two occupations—Signal Support Specialist (25U) and Human Resources Specialist (42A) that had previously been administered only the Army-wide measures.

We have also recently revised both the training and in-unit performance rating scales in an effort to improve their psychometric properties. For example, we have changed the format of the training MOS-specific rating scales to a 5-point relative performance rating rather than a 7-point absolute performance rating and greatly reduced the amount of reading required. The training Army-wide PRS have been similarly changed, and the number of dimensions rated has been reduced.

Data from the revised criterion measures will be incorporated into the next TOPS analysis data file.

Analyses

The semi-annual reports will continue to include basic psychometric, validation, and incremental validation analyses. As needed, we will examine the comparability of new TAPAS versions to prior forms before determining if the data can be combined for purposes of analysis. Analysis strategies also will be developed to handle data produced by substantially revised performance rating scales which started being administered in fall 2011. Finally, the plan is to conduct classification-oriented analyses annually.

The next set of TOPS evaluation analyses will be conducted based on a data file constructed in May 2012. The sample sizes for this next evaluation are expected to be considerably larger, thus supporting additional analyses yielding more generalizable results.

References

- Allen, M.T., Cheng, Y.A., Putka, D.J., Hunter, A., & White L. (2010). Analysis and findings. In D.J. Knapp & T.S. Heffner (Eds.). *Expanded enlistment eligibility metrics (EEEM): Recommendations on a non-cognitive screen for new soldier selection* (Technical Report 1267). Arlington, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.
- Burket, G. R. (1964). A study of reduced rank models for multiple prediction. *Psychometric Monograph* (No. 12).
- Campbell, J.P., & Knapp, D.J. (Eds.) (2001). *Exploring the limits in personnel selection and classification*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- Campbell, J.P., Hanson, M. A., & Oppler S. H. (2001). Modeling performance in a population of jobs. In J. P. Campbell & D. J. Knapp (Eds.), *Exploring the limits in personnel selection and classification*. Hillsdale, NJ: Erlbaum.
- Caramagno, J.P., Allen, M. T., & Ingerick, M. J. (2011). Evidence for the predictive validity of the TAPAS. In D.J. Knapp & T.S. Heffner (Eds.), *Tier One Performance Screen initial operational test and evaluation: 2010 Annual Report* (FR 11-40). Alexandria, VA: Human Resources Research Organization.
- Chernyshenko, O.S., & Stark, S. (October, 2007). *Criterion validity evidence for narrow temperament clusters: A meta-analysis of military studies*. Paper presented at the 49th annual conference of the International Military Testing Association. Gold Coast, AU.
- Chernyshenko, O.S., Stark, S., Drasgow, F., & Roberts, B.W. (2007). Constructing personality scales under the assumptions of an ideal point response process: Toward increasing the flexibility of personality measures. *Psychological Assessment*, 19, 88-106.
- Chernyshenko, O.S., Stark, S., Woo, S., & Conz, G. (2008, April). *Openness to Experience: Its facet structure, measurement and validity*. Paper presented at at the 23rd annual conference for the Society of Industrial and Organizational Psychologists. New Orleans, LA.
- Drasgow, F., Embretson, S. E., Kyllonen, P. C., & Schmitt, N. (2006). *Technical review of the Armed Services Vocational Aptitude Battery (ASVAB)* (FR-06-25). Alexandria, VA: Human Resources Research Organization.
- Ingerick, M., Diaz, T., & Putka, D. (2009). *Investigations into Army enlisted classification systems: Concurrent validation report* (Technical Report 1244). Arlington, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.
- Knapp, D. J., & Heffner, T. S. (Eds.) (2009). *Predicting Future Force Performance (Army Class): End of Training Longitudinal Validation* (Technical Report 1257). Arlington, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

- Knapp, D. J., & Heffner, T. S. (Eds.). (2010). *Expanded Enlistment Eligibility Metrics (EEEM): Recommendations on a non-cognitive screen for new soldier selection* (Technical Report 1267). Arlington, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.
- Knapp, D. J., & Heffner, T. S. (Eds.). (2011). *Tier One Performance Screen initial operational test and evaluation: 2010 annual report* (Technical Report 1296). Arlington, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.
- Knapp, D. J., & Heffner, T. S. (Eds.). (2012). *Tier One Performance Screen initial operational test and evaluation: 2011 interim report* (Technical Report 1306). Ft. Belvoir, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.
- Knapp, D. J., Heffner, T. S., & White, L. (Eds.). (2011). *Tier One Performance Screen initial operational test and evaluation: Early results* (Technical Report 1283). Arlington, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.
- Knapp, D. J., Owens, K. S., & Allen, M. T. (Eds.). (2011). *Validating future force performance measures (Army Class): First in-unit performance longitudinal validation* (Technical Report 1293). Arlington, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.
- Knapp, D. J., & Tremble, T. R. (Eds.). (2007). *Concurrent validation of experimental Army enlisted personnel selection and classification measures* (Technical Report 1205). Arlington, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.
- Moriarty, K. O., Campbell, R. C., Heffner, T. S., & Knapp, D. J. (2009). *Validating future force performance measures (Army Class): Reclassification test and criterion development* (Research Product 2009-11). Arlington, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.
- Putka, D. J., Le, H., McCloy, R. A., & Diaz, T. (2008). Ill-structured measurement designs in organizational research: Implications for estimating interrater reliability. *Journal of Applied Psychology*, 93, 959-981.
- Putka, D. J. & Van Iddekinge, C. H. (2007). Work Preferences Survey. In D. J. Knapp & T. R. Tremble (Eds.), *Concurrent validation of experimental Army enlisted personnel selection and classification measures* (Technical Report 1205). Arlington, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.
- Schmitt, N., & Ployhart, R. E. (1999). Estimates of cross-validity for stepwise regression and with predictor selection. *Journal of Applied Psychology*, 84(1), 50-57.
- Stark, S. E., Chernyshenko, O. S., & Drasgow, F. (2005). An IRT approach to constructing and scoring pairwise preference items involving stimuli on different dimensions: The multi-unidimensional pairwise preference model. *Applied Psychological Measurement*, 29, 184-201.

- Stark, S. E., Chernyshenko, O. S., & Drasgow, F. (2010a). Adaptive testing with the multi-dimensional pairwise preference model. Manuscript submitted for publication.
- Stark, S. E., Chernyshenko, O. S., & Drasgow, F. (2010b). Tailored adaptive personality assessment system (TAPAS-95s). In D. J. Knapp & T. S. Heffner (Eds.) *Expanded enlistment eligibility metrics (EEEM): Recommendations on a non-cognitive screen for new soldier selection* (Technical Report 1267). Arlington, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.
- Trippe, D. M., Caramagno, J. P., Allen, M. T., & Ingerick, M. J. (2011). Initial evidence for the predictive validity and classification potential of the TAPAS. In D. J. Knapp, T. S. Heffner, & L. White (Eds.), *Tier One Performance Screen initial operational test and evaluation: Early results* (TR 1283). Arlington, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.
- Trippe, D. M., Ford, F., Bynum, B., & Moriarty, K. O. (2012). Data file development. In D. J. Knapp & T. S. Heffner (Eds.), *Tier One Performance Screen initial operational test and evaluation: 2011 interim report* (FR 12-03). Alexandria, VA: Human Resources Research Organization.
- Van Iddekinge, C. H., Putka, D. J., & Sager, C. E. (2005). Attitudinal criteria. In D. J. Knapp & T. R. Tremble (Eds.), *Development of experimental Army enlisted personnel selection and classification tests and job performance criteria* (pp. 89-104) (Technical Report 1168). Arlington, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.
- White, L. A., & Young, M. C. (1998, August). *Development and validation of the Assessment of Individual Motivation (AIM)*. Paper presented at the annual meeting of the American Psychological Association, San Francisco, CA.

APPENDIX A

PREDICTOR MEASURE PSYCHOMETRIC PROPERTIES IN THE APPLICANT SAMPLE

Table A.1. Raw Mean and Standard Deviations for the TOPS Composites and TAPAS Scales by Version and Education Tier in the Applicant Sample

TAPAS Scale/ TOPS Composite	Tier 1 (n = 160,387)				Tier 2 (n = 7,898)				Tier 1 + 2 (Combined) (n = 168,285)			
	15D-Static (n = 19,150)		15D-CAT (n = 141,237)		15D-Static (n = 1,369)		15D-CAT (n = 6,592)		15D-Static (n = 20,519)		15D-CAT (n = 147,766)	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
<i>Individual TAPAS Scales</i>												
Achievement	.242	.494	.150	.482	.263	.502	.181	.472	.244	.495	.151	.482
Adjustment	.111	.588	-.016	.569	.144	.591	.050	.579	.114	.588	-.013	.570
Attention Seeking	-.252	.535	-.208	.531	-.261	.520	-.250	.521	-.252	.534	-.210	.530
Cooperation	-.060	.387	-.063	.372	-.068	.396	-.074	.373	-.061	.387	-.063	.372
Dominance	-.011	.583	.029	.588	-.056	.603	-.013	.599	-.014	.585	.027	.589
Even Tempered	.226	.482	.156	.476	.296	.478	.205	.471	.231	.482	.159	.476
Intellectual Efficiency	-.088	.583	-.029	.584	-.098	.558	.000	.560	-.088	.582	-.028	.583
Non-Delinquency	.128	.453	.099	.458	.056	.494	.025	.479	.123	.457	.095	.460
Optimism	.247	.494	.136	.459	.259	.490	.144	.448	.248	.494	.137	.458
Order	-.405	.567	-.423	.544	-.385	.555	-.420	.528	-.404	.566	-.422	.543
Physical Conditioning	-.004	.622	.049	.625	-.152	.579	-.092	.596	-.014	.620	.042	.625
Self Control	.088	.526	.061	.530	.143	.520	.124	.533	.092	.525	.064	.530
Selflessness	-.184	.442	-.200	.429	-.160	.442	-.186	.428	-.182	.442	-.199	.429
Sociability	-.167	.597	-.045	.592	-.184	.582	-.096	.592	-.168	.596	-.048	.592
Tolerance	-.249	.584	-.229	.568	-.248	.570	-.200	.555	-.249	.583	-.227	.567
<i>TOPS Composites</i>												
Can Do	.067	2.673	.002	2.736	.163	2.798	.081	2.749	.074	2.682	.005	2.736
Will Do	.124	2.374	.080	2.397	-.016	2.490	-.054	2.414	.115	2.382	.074	2.398

Note. Results are limited to the Applicant Sample (Non-prior service, Education Tier 1 and 2, AFQT Category IV and above) with valid TAPAS 15D-Static or 15D-CAT score data (June 2009-July 2011).

Table A.2. Correlations Among TOPS Composites and TAPAS Scale Scores by Education Tier in the Applicant Sample

TAPAS Scale/ TOPS Composite/	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
<i>Individual TAPAS Scales</i>																	
1. Achievement		.07	.06	.12	.30	.11	.22	.25	.18	.14	.15	.23	.10	.08	.11	.62	.59
2. Adjustment	.09		.13	.13	.06	.16	.15	.02	.28	-.08	.04	.08	.01	.14	.03	.24	.07
3. Attention Seeking	.05	.11		.10	.18	-.01	.08	-.08	.16	-.07	.11	-.09	-.07	.35	.07	.07	-.37
4. Cooperation	.11	.12	.06		.02	.25	.04	.19	.19	.00	-.02	.14	.19	.21	.15	.28	.18
5. Dominance	.33	.11	.20	.01		-.07	.21	.00	.13	.06	.18	.05	.00	.21	.03	.20	.10
6. Even Tempered	.11	.19	.00	.25	-.05		.07	.21	.20	-.03	-.08	.20	.12	.09	.13	.56	.50
7. Intellectual Efficiency	.26	.19	.08	.04	.25	.09		.03	.09	.04	.06	.19	.00	.02	.06	.49	.12
8. Non-Delinquency	.18	.00	-.13	.17	-.02	.17	.01		.13	.10	-.02	.24	.18	-.01	.12	.59	.63
9. Optimism	.19	.27	.17	.16	.17	.18	.10	.08		-.04	.08	.07	.10	.24	.12	.56	.18
10. Order	.15	-.08	-.10	.00	.05	-.03	.02	.10	-.02		.06	.18	.04	-.04	.02	.08	.14
11. Physical Conditioning	.16	.07	.12	-.01	.18	-.07	.05	-.02	.10	.03		-.03	-.03	.12	-.05	.07	.36
12. Self Control	.22	.07	-.11	.12	.05	.19	.18	.23	.06	.18	-.05		.08	-.08	.10	.33	.31
13. Selflessness	.08	-.02	-.08	.19	.01	.12	-.02	.13	.04	.05	-.04	.08		.08	.32	.18	.18
14. Sociability	.05	.11	.35	.18	.22	.04	.00	-.04	.23	-.04	.13	-.11	.07		.13	.15	-.03
15. Tolerance	.11	.02	.03	.15	.06	.13	.06	.06	.09	.03	-.06	.11	.32	.12		.19	.10
<i>TOPS Composites</i>																	
16. Can Do	.62	.27	.06	.25	.25	.55	.52	.52	.56	.08	.08	.31	.13	.09	.16		.72
17. Will Do	.57	.10	-.39	.18	.10	.49	.13	.59	.16	.14	.38	.29	.16	-.08	.09	.70	

Note. Results are limited to the Applicant Sample (Non-prior service, Education Tier 1 and 2, AFQT Category IV and above) with valid TAPAS 15D-CAT score data (June 2009-July 2011). Correlations below the diagonal are for Education Tier 1 applicants, $n = 141,237$. Correlations above the diagonal are for Education Tier 2 applicants, $n = 6,592$. Correlations in bold are statistically significant, $p < .01$ (two-tailed).

Table A.3. Correlations Among TOPS Composites and TAPAS Scale Scores in the Applicant Sample

TAPAS Scale/ TOPS Composite/	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
<i>Individual TAPAS Scales</i>																
1. Achievement																
2. Adjustment	.09															
3. Attention Seeking	.05	.11														
4. Cooperation	.11	.12	.06													
5. Dominance	.33	.10	.20	.01												
6. Even Tempered	.11	.19	.00	.25	-.05											
7. Intellectual Efficiency	.26	.18	.08	.04	.25	.09										
8. Non-Delinquency	.18	.00	-.13	.17	-.02	.18	.02									
9. Optimism	.19	.27	.17	.17	.17	.18	.10	.08								
10. Order	.15	-.08	-.09	.00	.05	-.03	.02	.10	-.02							
11. Physical Conditioning	.15	.07	.12	-.01	.18	-.07	.05	-.02	.10	.03						
12. Self Control	.22	.07	-.11	.12	.05	.19	.18	.23	.06	.18	-.05					
13. Selflessness	.09	-.02	-.08	.19	.01	.12	-.02	.13	.04	.05	-.04	.08				
14. Sociability	.05	.11	.35	.18	.22	.04	.00	-.04	.23	-.04	.13	-.11	.07			
15. Tolerance	.11	.02	.03	.15	.06	.13	.06	.06	.09	.03	-.06	.11	.32	.12		
<i>TOPS Composites</i>																
16. Can Do	.62	.27	.06	.26	.25	.56	.53	.52	.55	.08	.08	.32	.13	.10	.17	
17. Will Do	.57	.10	-.39	.19	.10	.49	.14	.59	.16	.14	.38	.29	.15	-.07	.08	.70

Note. Results are limited to the Applicant Sample (Non-prior service, Education Tier 1 and 2, AFQT Category IV and above) with valid TAPAS 15D-CAT score data (June 2009-July 2011), $n = 147,766$. Correlations in bold are statistically significant, $p < .01$ (two-tailed).

Table A.4. Correlations between TOPS Composites and TAPAS Scale Scores with AFQT by Education Tier in the Applicant Sample

TAPAS Scale/ TOPS Composite	AFQT		
	Tier 1	Tier 2	Tier 1 + 2 (Combined)
<i>N</i>	141,237	6,529	147,766
<i>Individual TAPAS Scales</i>			
Achievement	.10	.05	.09
Adjustment	.12	.11	.12
Attention Seeking	.11	.08	.11
Cooperation	.02	.02	.02
Dominance	.08	.00	.08
Even Tempered	.09	.09	.09
Intellectual Efficiency	.41	.38	.41
Non-Delinquency	-.01	.02	.00
Optimism	.02	.01	.02
Order	-.18	-.18	-.18
Physical Conditioning	.05	-.02	.05
Self Control	.00	.03	.00
Selflessness	-.07	-.05	-.06
Sociability	-.08	-.06	-.08
Tolerance	-.02	-.01	-.02
<i>TOPS Composites</i>			
Can Do	.22	.19	.22
Will Do	.05	.02	.05

Note. Results are limited to the Applicant Sample (Non-prior service, Education Tier 1 and 2, AFQT Category IV and above) with valid TAPAS 15D-CAT score data (June 2009-July 2011). Correlations in bold are statistically significant, $p < .01$ (two-tailed).

Table A.5. Descriptive Statistics for AFQT, ASVAB Subtests, and Aptitude Area (AA) Composites by Education Tier in the Applicant Sample

Subtest/Composite	Tier 1				Tier 2				Tier 1 + 2 (Combined)			
	<i>M</i>	<i>SD</i>	<i>Min</i>	<i>Max</i>	<i>M</i>	<i>SD</i>	<i>Min</i>	<i>Max</i>	<i>M</i>	<i>SD</i>	<i>Min</i>	<i>Max</i>
<i>AFQT</i>	56.83	23.25	10.00	99.00	53.49	19.24	10.00	99.00	56.67	23.08	10.00	99.00
<i>ASVAB Subtests</i>												
Arithmetic Reasoning (AR)	52.47	7.77	18.00	72.00	51.66	6.47	24.00	72.00	52.43	7.71	18.00	72.00
Assembling Objects (AO)	55.00	7.89	25.00	70.00	54.48	7.68	26.00	69.00	54.97	7.88	25.00	70.00
Auto & Shop Information (AS)	49.72	9.41	19.00	86.00	51.97	9.02	24.00	81.00	49.83	9.40	19.00	86.00
Electronics Information (EI)	51.79	9.15	16.00	84.00	52.84	8.06	18.00	83.00	51.84	9.10	16.00	84.00
General Science (GS)	51.56	8.47	19.00	76.00	51.34	7.14	23.00	74.00	51.54	8.41	19.00	76.00
Math Knowledge (MK)	53.55	6.99	24.00	73.00	49.74	5.77	28.00	73.00	53.37	6.98	24.00	73.00
Mechanical Comprehension (MC)	53.24	8.50	14.00	82.00	53.59	7.66	23.00	80.00	53.26	8.46	14.00	82.00
Paragraph Comprehension (PC)	52.70	7.13	23.00	69.00	52.91	6.23	21.00	69.00	52.71	7.09	21.00	69.00
Word Knowledge (WK)	51.16	8.15	16.00	76.00	51.55	6.78	22.00	76.00	51.18	8.08	16.00	76.00
<i>Aptitude Area (AA) Composites</i>												
Clerical (CL)	105.69	14.09	34.72	151.97	103.36	11.07	56.07	145.20	105.57	13.96	34.72	151.97
Combat (CO)	105.42	15.03	28.95	159.85	104.06	12.19	55.95	154.19	105.36	14.90	28.95	159.85
Electronics (EL)	105.18	15.03	28.62	159.59	104.34	12.14	56.71	153.39	105.14	14.91	28.62	159.59
Field Artillery (FA)	105.59	14.95	28.46	159.14	104.16	12.08	56.97	153.94	105.52	14.83	28.46	159.14
General Maintenance (GM)	104.87	15.49	28.14	160.64	104.34	12.73	56.56	155.24	104.84	15.37	28.14	160.64
Mechanical Maintenance (MM)	103.89	16.46	24.59	164.53	105.02	14.20	55.52	158.33	103.94	16.36	24.59	164.53
Operators & Food Service (OF)	104.84	15.46	27.17	160.54	104.67	12.71	57.53	154.42	104.83	15.34	27.17	160.54
Signal Communications (SC)	105.62	14.65	29.38	158.57	104.01	11.69	55.69	152.66	105.55	14.53	29.38	158.57
Skilled Technical (ST)	105.47	14.67	31.54	156.85	104.18	11.68	57.36	149.73	105.41	14.54	31.54	156.85
<i>N</i>												
AFQT	180,646				9,180				189,826			
ASVAB Subtests	176,902 - 179,662				8,990 - 9,163				185,892 - 188,825			
AA Composites	179,664				9,162				188,826			

Note. Results are limited to the Applicant Sample (non-prior service, Education Tier 1 and 2, AFQT Category IV and above) (June 2009-July 2011).

Table A.6. Correlations Among AFQT, ASVAB Subtests, and AA Composite Scores by Education Tier in the Applicant Sample

Subtest/Composite	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1. AFQT		.77	.38	.28	.53	.68	.65	.57	.76	.77	.95	.83	.86	.85	.80	.66	.81	.90	.91
<i>ASVAB Subtests</i>																			
2. Arithmetic Reasoning (AR)	.82		.41	.24	.38	.44	.61	.50	.46	.37	.87	.73	.74	.76	.71	.58	.73	.77	.78
3. Assembling Objects (AO)	.45	.48		.23	.31	.29	.31	.49	.28	.22	.44	.47	.45	.48	.45	.41	.46	.46	.46
4. Auto & Shop Info (AS)	.38	.32	.27		.65	.45	-.04	.58	.26	.33	.37	.67	.66	.64	.73	.87	.72	.57	.55
5. Electronics Information (EI)	.60	.48	.36	.69		.64	.18	.64	.46	.54	.59	.77	.79	.75	.81	.82	.78	.76	.72
6. General Science (GS)	.74	.55	.37	.51	.69		.31	.60	.57	.68	.67	.76	.76	.75	.76	.69	.75	.74	.77
7. Math Knowledge (MK)	.72	.70	.39	.07	.30	.44		.29	.32	.22	.71	.57	.54	.59	.49	.32	.47	.62	.59
8. Mechanical Comp (MC)	.66	.61	.53	.62	.69	.67	.42		.45	.46	.65	.83	.80	.83	.82	.82	.83	.77	.79
9. Paragraph Comp (PC)	.81	.56	.36	.36	.54	.66	.43	.56		.63	.70	.62	.66	.63	.60	.52	.62	.67	.70
10. Word Knowledge (WK)	.81	.49	.29	.41	.60	.73	.36	.55	.71		.68	.64	.69	.64	.62	.56	.65	.69	.73
<i>Aptitude Area (AA) Composites</i>																			
11. Clerical (CL)	.96	.90	.50	.44	.65	.74	.78	.73	.76	.75		.91	.93	.93	.88	.75	.89	.96	.96
12. Combat (CO)	.88	.80	.52	.68	.79	.81	.67	.86	.71	.72	.94		.99	1.00	.99	.94	.99	.99	.98
13. Electronics (EL)	.90	.81	.51	.67	.81	.81	.65	.84	.74	.76	.95	1.00		.99	.99	.94	.99	.99	.99
14. Field Artillery (FA)	.89	.82	.53	.65	.77	.80	.69	.86	.71	.71	.95	1.00	1.00		.99	.93	.99	.99	.99
15. General Maintenance (GM)	.86	.79	.51	.73	.83	.81	.61	.85	.69	.70	.92	1.00	.99	.99		.97	1.00	.97	.96
16. Mechanical Maintenance (MM)	.75	.68	.47	.86	.85	.75	.46	.86	.63	.66	.81	.95	.95	.94	.97		.97	.89	.88
17. Operators & Food (OF)	.87	.80	.51	.73	.81	.80	.59	.86	.71	.73	.92	.99	.99	.99	1.00	.97		.97	.97
18. Signal Communications (SC)	.93	.83	.51	.60	.79	.79	.71	.82	.75	.75	.97	.99	1.00	.99	.98	.92	.98		.99
19. Skilled Technical (ST)	.94	.83	.51	.59	.75	.82	.69	.83	.77	.78	.98	.99	.99	.99	.97	.91	.98	1.00	

Note. Results are limited to the Applicant Sample (Non-prior service, Education Tier 1 and 2, AFQT Category IV and above) (June 2009-July 2011). Correlations below the diagonal are for Education Tier 1 applicants, $n = 176,902 - 180,646$. Correlations above the diagonal are for Education Tier 2 applicants, $n = 8,990 - 9,180$. All correlations are statistically significant, $p < .01$ (one-tailed).

Table A.7. Correlations Among AFQT, ASVAB Subtests, and AA Composite Scores in the Applicant Sample

Subtest/Composite	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1. AFQT																		
<i>ASVAB Subtests</i>																		
2. Arithmetic Reasoning (AR)	.82																	
3. Assembling Objects (AO)	.45	.48																
4. Auto & Shop Info (AS)	.37	.31	.27															
5. Electronics Information (EI)	.60	.48	.36	.69														
6. General Science (GS)	.74	.55	.36	.51	.69													
7. Math Knowledge (MK)	.72	.69	.39	.06	.29	.43												
8. Mechanical Comp (MC)	.66	.60	.53	.62	.69	.67	.41											
9. Paragraph Comp (PC)	.80	.56	.35	.36	.54	.65	.42	.56										
10. Word Knowledge (WK)	.81	.49	.29	.41	.59	.73	.35	.55	.71									
<i>Aptitude Area (AA) Composites</i>																		
11. Clerical (CL)	.96	.90	.50	.43	.64	.74	.78	.72	.76	.74								
12. Combat (CO)	.88	.79	.52	.67	.79	.81	.67	.86	.71	.71	.94							
13. Electronics (EL)	.90	.80	.51	.67	.81	.81	.64	.83	.74	.75	.95	1.00						
14. Field Artillery (FA)	.89	.82	.53	.65	.77	.80	.68	.86	.71	.71	.95	1.00	.99					
15. General Maintenance (GM)	.85	.78	.51	.73	.83	.81	.60	.85	.69	.70	.91	.99	.99	.99				
16. Mechanical Maintenance (MM)	.74	.67	.47	.86	.85	.75	.45	.86	.63	.65	.81	.95	.95	.94	.97			
17. Operators & Food (OF)	.87	.79	.51	.73	.81	.80	.59	.86	.71	.72	.92	.99	.99	.99	1.00	.97		
18. Signal Communications (SC)	.92	.83	.51	.60	.78	.79	.71	.81	.74	.75	.97	.99	1.00	.99	.98	.92	.98	
19. Skilled Technical (ST)	.94	.83	.51	.58	.75	.81	.68	.82	.77	.78	.98	.99	.99	.99	.97	.91	.98	1.00

Note. Results are limited to the TOPS Applicant Sample (Non-prior service, Education Tier 1 and 2, AFQT Category IV and above) (June 2009-July 2011), $n = 185,891 - 189,826$. All correlations are statistically significant, $p < .01$ (one-tailed).

Table A.8. Group Means on the AFQT and TAPAS Scales by Gender and Race-Ethnicity in the Applicant Sample

Measure/ Scale/Composite	Gender		Race-Ethnicity		
	Female	Male	Black	Hispanic	White, Non- Hispanic
<i>AFQT</i>	52.37	58.20	45.61	47.21	62.70
<i>Individual TAPAS Scales</i>					
Achievement	.03	-.01	-.05	-.06	.04
Adjustment	-.23	.06	-.08	-.11	.07
Attention Seeking	-.08	.02	-.05	-.03	.03
Cooperation	-.03	.01	-.01	-.04	.00
Dominance	-.10	.02	.03	.01	.01
Even Tempered	-.09	.02	.00	-.08	.03
Intellectual Efficiency	-.17	.04	-.09	-.12	.07
Non-Delinquency	.12	-.03	.05	-.05	.01
Optimism	-.03	.00	.03	.00	.01
Order	.13	-.03	.19	.13	-.10
Physical Conditioning	-.33	.08	-.17	-.07	.07
Self Control	-.01	.00	.16	.06	-.06
Selflessness	.30	-.08	.10	.04	-.05
Sociability	.01	.00	-.04	.00	.01
Tolerance	.27	-.07	.18	.22	-.13
<i>TOPS Composites</i>					
Can Do	-.05	.01	-.02	-.12	.06
Will Do	-.07	.01	-.05	-.10	.05
<i>N</i>					
AFQT	31,653	127,143	24,382	21,934	89,967
TAPAS/TOPS Composites	29,356	117,882	22,396	20,186	83,349

Note. Results are limited to the Applicant Sample (Non-prior service, Education Tier 1 and 2, AFQT Category IV and above) with valid TAPAS 15D-CAT score data (June 2009-July 2011).

Table A.9. Standardized Group Mean Score Differences on the AFQT and TAPAS in the Applicant Sample

Measure/ Scale/Composite	Female- Male	Black-White, Non- Hispanic	Hispanic- White, Non- Hispanic
<i>AFQT</i>	-.25	-.77	-.70
<i>Individual TAPAS Scales</i>			
Achievement	.04	-.09	-.10
Adjustment	-.29	-.14	-.18
Attention Seeking	-.11	-.08	-.06
Cooperation	-.04	-.01	-.04
Dominance	-.12	.02	.00
Even Tempered	-.11	-.03	-.11
Intellectual Efficiency	-.21	-.16	-.18
Non-Delinquency	.15	.04	-.07
Optimism	-.04	.01	-.01
Order	.16	.29	.23
Physical Conditioning	-.41	-.23	-.14
Self-Control	-.01	.21	.12
Selflessness	.39	.15	.08
Sociability	.01	-.05	-.01
Tolerance	.34	.30	.34
<i>Min d</i>	<i>-.41</i>	<i>-.23</i>	<i>-.18</i>
<i>Max d</i>	<i>.39</i>	<i>.30</i>	<i>.34</i>
<i>Average Absolute d</i>	<i>.16</i>	<i>.12</i>	<i>.11</i>
<i>TOPS Composites</i>			
Can Do	-.06	-.08	-.18
Will Do	-.09	-.10	-.15
<i>N</i>			
AFQT	158,796	114,349	111,890
TAPAS/TOPS Composites	147,238	105,745	103,535

Note. Results are limited to the Applicant Sample (Non-prior service, Education Tier 1 and 2, AFQT Category IV and above) with valid TAPAS 15D-CAT score data (June 2009-July 2011). A negative standardized mean difference (or Cohen's *d*) indicates that the minority (Female, Black, Hispanic) group's mean is lower than the majority (Male; White, Non-Hispanic) group's mean, on average. *d*'s in bold are statistically significant, $p < .01$ (two-tailed), based on independent samples *t*-test analyses of group mean differences.

APPENDIX B

CRITERION PSYCHOMETRIC PROPERTIES IN THE FULL IMT SAMPLE

Table B.1. Descriptive Statistics and Internal Consistency Reliability Estimates for WTBD and MOS-Specific Job Knowledge Tests (JKTs) in the Full IMT Sample

Domain / MOS	<i>n</i>	<i>M</i>	<i>SD</i>	<i>Min</i>	<i>Max</i>	<i>r</i> _{WTBD}	α
<i>MOS-Specific</i>							
11B/C/X + 18X	10,499	61.7	10.0	20.9	88.4	.54	.76
19K	223	61.0	9.8	30.0	82.0	.55	.66
31B	4,960	69.4	8.5	35.0	93.2	.50	.77
68W	6,555	74.0	10.5	25.0	96.7	.50	.87
88M	4,402	63.5	11.1	30.6	94.4	.54	.77
91B	1,051	57.2	13.6	23.7	90.7	.47	.90
<i>All MOS Combined</i>	27,690	66.1	11.5	20.9	97.0	.54	--
<i>WTBD (Army-Wide)</i>							
IMT	33,865	64.8	12.9	6.0	100.0	--	.65

Note. Means, SDs, Min, and Max are based on percent correct; α = coefficient alpha. WTBD = Warrior Tasks and Battle Drills. *r*_{WTBD} = correlation with WTBD JKT scores. All correlations are statistically significant ($p < .01$, one-tailed).

Table B.2. Descriptive Statistics and Reliability Estimates for Army-Wide and MOS-Specific Performance Rating Scales (PRS) in the Full IMT Sample

Domain / PRS	<i>n</i>	<i>M</i>	<i>SD</i>	<i>Min</i>	<i>Max</i>	α	<i>IRR</i>
<i>Army-Wide</i>							
Can Do ^a	11,775	4.97	1.13	1.00	7.00	.88	.08
Commitment & Adjustment	12,077	5.09	1.22	1.00	7.00	--	.17
Effort & Personal Discipline ^a	12,104	4.99	1.16	1.00	7.00		
Physical Fitness & Bearing	12,005	4.90	1.22	1.00	7.00	--	.23
Working with Others ^a	12,048	4.90	1.17	1.00	7.00	.84	.16
Overall Performance	11,911	3.51	.84	1.00	5.00	--	.32
Army-Wide ^a	12,105	4.96	1.05	1.00	7.00	.95	.19
<i>MOS-Specific</i>							
11B/C/X + 18X	3,964	4.86	.93	1.00	7.00	.95	.20
19K	158	5.27	.56	2.71	6.86	.85	.38
31B	1,794	5.03	.98	1.00	7.00	.95	.12
68W	3,096	4.49	.91	1.00	7.00	.93	.02
88M	674	4.80	.94	2.00	7.00	.94	.00
91B	209	4.53	1.56	1.00	7.00	.96	.20
<i>All MOS Combined</i>	9,895	4.77	.97	1.00	7.00	--	--

Note. Ratings on PRS range from 1 and 7, except for the Overall Performance PRS, which ranges from 1 to 5. PRS ratings from supervisors with a familiarity rating of 1 ("I have had little opportunity to observe this Soldier") were excluded from analyses. α = coefficient alpha. IRR = Interrater reliability, estimated using G(q,k) (Putka, Le, McCloy, & Diaz, 2008). IRR estimates were not estimated if 30 or fewer Soldiers were rated by more than one supervisor.

^a Ratings composite comprises two or more Army-wide PRS.

Table B.3. Correlations Among Army-Wide and MOS-Specific PRS in the Full IMT Sample

Domain/PRS	1	2	3	4	5	6	7
<i>Army-Wide</i>							
1. Can Do ^a							
2. Commitment & Adjustment	.76						
3. Effort & Personal Discipline ^a	.74	.82					
4. Physical Fitness & Bearing	.66	.70	.74				
5. Working with Others ^a	.78	.78	.79	.68			
6. Overall Performance	.56	.57	.61	.57	.60		
7. Army-Wide ^a	.89	.90	.92	.82	.92	.65	
<i>MOS-Specific</i>							
8. 11B/C/X + 18X	.76	.67	.67	.61	.68	.55	.76
9. 19K	.75	.71	.69	.65	.51	.61	.77
10. 31B	.75	.66	.65	.56	.66	.55	.75
11. 68W	.59	.51	.48	.40	.52	.35	.56
12. 88M	.73	.66	.60	.60	.65	.54	.73
13. 91B	.93	.81	.82	.75	.81	.65	.90
14. All MOS Combined	.72	.64	.62	.56	.63	.49	.72

Note. Army-wide PRS, $n = 11,596 - 12,105$. MOS-specific PRS, 11B, $n = 3,599 - 3,603$; 19K, $n = 149$; 31B, $n = 1,661 - 1,675$; 68W, $n = 2,040 - 2,270$; 88M, $n = 609 - 626$; 91B, $n = 182 - 201$. Ratings on PRS range from 1 and 7, except for the Overall Performance PRS, which ranges from 1 to 5. PRS ratings from supervisors with a familiarity rating of 1 ("I have had little opportunity to observe this Soldier") were excluded from analyses. All correlations are statistically significant ($p < .01$, one-tailed).

^aRatings composite comprises two or more Army-wide PRS.

Table B.4. Descriptive Statistics and Internal Consistency Reliability Estimates for the Army Life Questionnaire (ALQ) in the Full IMT Sample

Domain / Scale	n	M	SD	Min	Max	α
<i>Retention</i>						
Affective Commitment	35,337	3.87	.68	1.00	5.00	.86
Army Life Adjustment	35,337	4.08	.66	1.00	5.00	.86
Army Fit	35,337	4.06	.60	1.00	5.00	.86
Attrition Cognitions	35,337	1.53	.61	1.00	5.00	.77
MOS Fit	35,337	3.77	.85	1.00	5.00	.92
<i>Achievement/Performance</i>						
Disciplinary Incidents	23,158	.26	.61	0.00	7.00	--
Last APFT Score	34,919	250.04	31.66	100.00	300.00	--
Training Achievements	35,301	.41	.61	0.00	2.00	--
Training Restarts	35,335	.40	.64	0.00	4.00	--

Note. α = coefficient alpha.

Table B.5. Correlations Among ALQ Scales in the Full IMT Sample

Domain / Scale	1	2	3	4	5	6	7	8
<i>Retention</i>								
1. Affective Commitment								
2. Army Life Adjustment	.46							
3. Army Fit	.84	.61						
4. Attrition Cognitions	-.63	-.54	-.69					
5. MOS Fit	.48	.35	.49	-.42				
<i>Achievement/Performance</i>								
6. Disciplinary Incidents	-.08	-.17	-.11	.12	-.08			
7. Last APFT Score	.05	.24	.11	-.12	.08	-.14		
8. Training Achievement	.06	.13	.07	-.04	.05	-.07	.24	
9. Training Restarts	-.07	-.20	-.10	.12	-.09	.20	-.28	-.12

Note. $n = 22,907 - 35,337$. All correlations are statistically significant ($p < .01$, two-tailed).

Table B.6. Correlations Among the IMT Army-Wide and MOS-Specific PRS in the Validation Sample

Domain/PRS	1	2	3	4	5	6	7
<i>Army-Wide</i>							
1. Can Do ^a							
2. Commitment & Adjustment	.78						
3. Effort & Personal Discipline ^a	.75	.81					
4. Physical Fitness & Bearing	.64	.68	.74				
5. Working with Others ^a	.77	.77	.78	.66			
6. Overall Performance	.56	.56	.59	.53	.57		
7. Army-Wide ^a	.90	.89	.93	.81	.91	.63	
<i>MOS-Specific</i>							
8. 11B/C/X + 18X	.72	.62	.65	.59	.65	.53	.74
9. 19K	.84	.71	.65	.72	.42	.65	.78
10. 31B	.75	.67	.63	.50	.65	.53	.74
11. 68W	.68	.68	.65	.52	.67	.38	.72
12. 88M	.73	.68	.68	.67	.70	.61	.77
13. 91B	--	--	--	--	--	--	--
14. All MOS Combined	.73	.66	.66	.56	.65	.49	.74

Note. Army-wide PRS, $n = 2,895-2,935$. MOS-specific PRS, 11B, $n = 872-875$; 19K, $n = 56$; 31B, $n = 540-543$; 68W, $n = 462-468$; 88M, $n = 113$; All MOS Combined, $n = 2,077 - 2,086$. Ratings on PRS range from 1 and 7, except for the Overall Performance PRS, which ranges from 1 to 5. PRS ratings from supervisors with a familiarity rating of 1 ("I have had little opportunity to observe this Soldier") were excluded from analyses. All correlations are statistically significant ($p < .01$, one-tailed).

^a Ratings composite comprises two or more Army-wide PRS.

Table B.7. Correlations Among the IMT ALQ Scales in the Validation Sample

Domain / Scale	1	2	3	4	5	6	7	8
<i>Retention</i>								
1. Affective Commitment								
2. Army Life Adjustment	.46							
3. Army Fit	.84	.61						
4. Attrition Cognitions	-.63	-.54	-.69					
5. MOS Fit	.48	.36	.48	-.41				
<i>Achievement/Performance</i>								
6. Disciplinary Incidents	-.07	-.16	-.09	.10	-.08			
7. Last APFT Score	.03	.21	.08	-.10	.08	-.14		
8. Training Achievement	.05	.13	.05	-.03	.06	-.06	.23	
9. Training Restarts	-.06	-.20	-.09	.11	-.10	.20	-.26	-.11

Note. $n = 7,410 - 9,143$. All correlations are statistically significant ($p < .01$, two-tailed).

Table B.8. Correlations between the IMT JKTs and PRS in the Validation Sample

Domain / PRS	JKT							
	All MOS Combined	11B	19K	31B	68W	88M	91B	WTBD
<i>Army-Wide</i>								
Can Do	.03	.06	.29	.08	-.02	.07	--	.08
Commitment and Adjustment	.04	.06	.25	.06	.00	.18	--	.07
Effort & Personal Discipline	.06	.07	.26	.11	.02	.21	--	.10
Physical Fitness and Bearing	.03	.02	.34	.08	.01	.21	--	.07
Working with Others	.05	.06	.20	.10	.02	.19	--	.08
Overall Performance	.02	.01	.29	.05	-.01	.11	--	.07
Army-Wide	.05	.06	.31	.10	.01	.19	--	.10
<i>MOS-Specific</i>								
11B	.08	.08	--	--	--	--	--	.11
19K	--	--	--	--	--	--	--	.37
31B	.06	--	--	.06	--	--	--	.14
68W	.04	--	--	--	.04	--	--	.06
88M	.07	--	--	--	--	.07	--	.04
91B	--	--	--	--	--	--	--	--
All MOS Combined	.06	.08	--	.06	.04	.07	--	.11

Note. WTBD = Warrior Tasks and Battle Drills. Army-wide PRS, All MOS Combined, $n = 2,399 - 2,420$; 11B, $n = 848 - 851$; 19K, $n = 61$; 31B, $n = 597 - 599$; 68W, $n = 765 - 783$; 88M, $n = 95 - 99$; WTBD, $n = 2,737 - 2,758$. MOS-specific PRS, All MOS Combined, $n = 90 - 1,925$; 11B, $n = 750$; 31B, $n = 523$; 68W, $n = 495$; 88M, $n = 90$. Correlations in bold are statistically significant ($p < .01$, two-tailed).

Table B.9. Correlations between the IMT JKTs and ALQ Scales in the Validation Sample

Domain / Scale	JKT							
	All MOS							
	Combined	11B	19K	31B	68W	88M	91B	WTBD
<i>Retention</i>								
Affective Commitment	.07	.09	.08	.06	.05	.02	.26	.08
Army Life Adjustment	.12	.12	-.05	.13	.14	.08	.18	.13
Army Fit	.13	.15	.13	.07	.14	.07	.28	.14
Attrition Cognitions	-.16	-.17	-.05	-.13	-.19	-.11	-.26	-.18
MOS Fit	.10	.09	.01	.04	.18	.03	.30	.13
<i>Achievement/Performance</i>								
Disciplinary Incidents	-.04	-.01	.14	-.09	-.08	-.02	-.07	-.04
Last APFT Score	.04	.03	.07	-.04	.02	.00	.03	.09
Training Achievement	-.09	-.15	-.24	-.03	.02	-.12	-.20	-.09
Training Restarts	-.07	-.07	.06	-.09	-.04	-.11	-.01	-.09

Note. WTBD = Warrior Tasks and Battle Drills. All MOS Combined, $n = 5,999-7,329$; 11B, $n = 2,959-2,980$; 19K, $n = 81-83$; 31B, $n = 1,013-1,398$; 68W, $n = 1,011-1,537$; 88M, $n = 790-1,120$; 91B, $n = 122-211$; WTBD, $n = 7,044-8,651$. Correlations in bold are statistically significant ($p < .01$, two-tailed).

Table B.10. Correlations between the IMT ALQ Scales and PRS in the Validation Sample

Domain / PRS	ALQ Scale								
	AFF COM	LIFE ADJ	Army Fit	ATT COG	MOS Fit	DSC INC	LAST APFT	TRN ACH	TRN RST
<i>Army-Wide</i>									
Can Do	.05	.07	.07	-.05	.07	-.10	.10	.06	-.05
Commitment and Adjustment	.06	.07	.07	-.05	.07	-.13	.11	.08	-.06
Effort & Personal Discipline	.06	.08	.08	-.05	.08	-.14	.11	.08	-.04
Physical Fitness and Bearing	.04	.10	.06	-.06	.07	-.11	.24	.11	-.10
Working with Others	.04	.04	.06	-.05	.05	-.09	.09	.05	-.05
Overall Performance	.05	.12	.07	-.08	.08	-.14	.18	.15	-.11
Army-Wide	.06	.08	.07	-.06	.08	-.12	.13	.08	-.06
<i>MOS-Specific</i>									
11B	.04	.08	.08	-.12	.11	-.07	.09	.02	-.03
19K	.07	.17	.12	-.16	.09	.00	.08	.05	-.19
31B	.12	.20	.16	-.12	.12	-.20	.06	.12	-.10
68W	.03	.07	.05	-.09	.02	-.04	-.02	.01	.04
88M	.01	-.14	.03	.10	-.08	-.05	.20	-.05	-.03
91B	--	--	--	--	--	--	--	--	--
All MOS Combined	.08	.12	.12	-.11	.08	-.10	.08	.05	-.05

Note. AFFCOM=Affective Commitment; LIFEADJ=Army Life Adjustment; ATTCOG=Attrition Cognitions; DSCINC=Disciplinary Incidents; LAST APFT=Last APFT Score; TRNACH=Training Achievements; TRNRST=Training Restarts. Army-wide PRS, $n = 2,374-2,860$. MOS-specific PRS, All MOS Combined, $n = 1,893-2,268$; 11B, $n = 906-913$; 19K, $n = 55-56$; 31B, $n = 457-576$; 68W, $n = 374-574$; 88M, $n = 86-118$. Correlations in bold are statistically significant ($p < .01$, two-tailed).

Table B.11. Correlations between the IMT JKTs and Administrative Criteria in the Validation Sample

Domain/Measure	JKT							
	All MOS							
	Combined	11B	19K	31B	68W	88M	91B	WTBD
<i>Attrition</i>								
6-Month Cumulative	-.01	-.02	--	.03	--	--	--	.01
9-Month Cumulative	-.04	-.04	--	-.07	.02	-.10	-.06	-.03
12-Month Cumulative	-.03	-.01	--	-.10	-.12	--	-.07	-.01
15-Month Cumulative	-.08	-.05	--	-.19	-.25	.00	--	-.05
18-Month Cumulative	-.17	-.15	--	-.31	-.18	--	--	-.10
<i>IMT Restarts</i>								
Restarted at Least Once During IMT	.01	.03	--	.05	-.03	.00	-.06	.01
Academic or Other Pejorative Restart	.02	.03	--	.02	.03	.00	-.06	.02
Academic Restart	.01	.03	--	.07	-.03	-.01	-.07	.01
<i>Final AIT School Grades</i>								
Overall Average (Unstandardized)	.31	--	--	--	.36	--	--	.33
Overall Average (Standardized)	.34	--	--	--	.34	--	--	.41

Note. WTBD = Warrior Tasks and Battle Drills. Attrition, All MOS Combined, $n = 619 - 4,164$; 11B, $n = 249 - 2,253$; 31B, $n = 68 - 570$; 68W, $n = 220 - 854$; 88M, $n = 151 - 264$; 91B, $n = 78 - 112$; WTBD, $n = 745 - 4,892$.

IMT Restarts, All MOS Combined, $n = 4,915 - 4,995$; 11B, $n = 2,198 - 2,206$; 31B, $n = 916 - 923$; 68W, $n = 889 - 923$; 88M, $n = 656 - 692$; 91B, $n = 201 - 204$; WTBD, $n = 5,856 - 5,950$.

Final AIT School Grade, All MOS Combined, $n = 82 - 83$; 68W, $n = 74 - 75$; WTBD, $n = 91 - 92$. Correlations in bold are statistically significant ($p < .01$, two-tailed).

Table B.12. Correlations between the IMT ALQ Scales and Administrative Criteria in the Validation Sample

Domain/Measure	ALQ Scale								
	AFF COM	LIFE ADJ	Army Fit	ATT COG	MOS Fit	DSC INC	LAST APFT	TRN ACH	TRN RST
<i>Attrition</i>									
6-Month Cumulative	-.03	-.08	-.05	.14	-.04	.06	-.10	-.02	.08
9-Month Cumulative	-.07	-.08	-.09	.20	-.06	.08	-.07	-.02	.06
12-Month Cumulative	-.03	-.08	-.06	.17	-.06	.09	-.05	-.02	.04
15-Month Cumulative	-.02	-.07	-.04	.15	-.02	.13	-.04	-.05	.00
18-Month Cumulative	-.05	-.07	-.10	.15	-.04	.17	.01	-.02	-.02
<i>IMT Restarts</i>									
Restarted at Least Once During IMT	.00	.01	.00	-.04	.02	-.07	.07	-.02	-.20
Academic or Other Pejorative Restart	-.01	.03	.00	-.05	.02	-.08	.07	.01	-.22
Academic Restart	.00	.00	.00	-.02	.01	-.06	.04	-.03	-.20
<i>Final AIT School Grades</i>									
Overall Average (Unstandardized)	.03	-.04	.07	-.04	.07	--	-.19	.12	.06
Overall Average (Standardized)	-.03	.04	.03	-.04	-.07	--	-.20	.04	-.04

Note. AFFCOM=Affective Commitment; LIFEADJ=Army Life Adjustment; ATTCOG=Attrition Cognitions; DSCINC=Disciplinary Incidents; LAST APFT=Last APFT Score; TRNACH=Training Achievements; TRNRST=Training Restarts. Attrition, $n = 336 - 5,155$. IMT Restarts, $n = 4,938 - 6,216$. Final AIT School Grade, $n = 93 - 95$. Correlations in bold are statistically significant ($p < .01$, two-tailed).

Table B.13. Correlations between the IMT PRS and Administrative Criteria in the Validation Sample

Domain/PRS	Attrition					IMT Restarts		
	6-Month	9-Month	12-Month	15-Month	18-Month	IMT Restart	PEJ Restart	ACAD Restart
<i>Army-Wide</i>								
Can Do	-.05	-.06	-.06	-.10	-.07	.02	.04	.00
Commitment & Adjustment	-.07	-.09	-.07	-.07	-.06	.05	.06	.03
Effort & Personal Discipline	-.07	-.04	-.01	-.06	-.05	.01	.02	-.01
Physical Fitness & Bearing	-.13	-.12	-.10	-.06	-.04	.03	.05	.00
Working with Others	-.05	-.04	-.01	-.08	-.05	.02	.03	.01
Overall Performance	-.08	-.07	-.08	-.09	-.09	.01	.04	-.02
Army-Wide	-.08	-.07	-.05	-.09	-.06	.03	.04	.01
<i>MOS-Specific</i>								
11B	.07	.07	.00	-.03	-.03	-.05	-.04	-.04
19K	--	--	--	--	--	--	--	--
31B	-.06	-.13	-.03	--	--	.05	.05	.01
68W	--	.00	.01	.02	-.08	-.07	.05	-.07
88M	--	--	--	--	--	-.21	.04	-.21
91B	--	--	--	--	--	--	--	--
<i>All MOS Combined</i>	.03	.02	.00	-.01	-.04	-.02	.02	-.03

Note. IMT Restart = Restarted at Least Once During IMT. PEJ Restart = Restarted at Least Once for Academic or Other Pejorative Reasons. ACAD Restart = Restarted at Least Once for Academic Reasons. Army-wide PRS, Attrition, $n = 276$ -1,735; IMT Restarts, $n = 2,111$ -2,163. MOS-specific PRS, Attrition, $n = 50$ -1,403; IMT Restarts, $n = 118$ -1,738. Final AIT Grade not included because samples sizes are < 50 . Correlations in bold are statistically significant ($p < .01$, two-tailed).

APPENDIX C

CRITERION PSYCHOMETRIC PROPERTIES IN THE FULL IN-UNIT SAMPLE

Table C.1. Descriptive Statistics and Internal Consistency Reliability Estimates for WTBD and MOS-Specific Job Knowledge Tests (JKTs) in the Full In-Unit Sample

Domain/Setting/MOS	<i>n</i>	<i>M</i>	<i>SD</i>	<i>Min</i>	<i>Max</i>	<i>r</i> _{WTBD}	<i>α</i>
<i>MOS-Specific (In-Unit)</i>							
11B/C/X + 18X	369	64.6	10.7	26.8	84.5	.58	.74
19K	--	--	--	--	--	--	--
31B	--	--	--	--	--	--	--
68W	76	75.7	8.0	50.9	90.6	.36	.57
88M	94	62.0	10.4	40.4	87.2	.67	.81
91B	80	55.7	10.6	35.1	80.7	.36	.75
<i>All MOS Combined</i>	696	64.6	11.7	26.8	90.6	.57	--
<i>WTBD (Army-Wide)</i>							
In-Unit	1,843	67.9	13.8	15.0	100.0	--	.62

Note. Means, SDs, Min, and Max are based on percent correct; α = coefficient alpha. WTBD = Warrior Tasks and Battle Drills. r_{WTBD} = correlation with WTBD JKT scores. All correlations are statistically significant ($p < .05$, one-tailed). Statistics based on fewer than 50 cases are not reported.

Table C.2. Descriptive Statistics and Reliability Estimates for the Performance Rating Scales (PRS) in the Full In-Unit Sample

Domain/Setting/PRS	<i>n</i>	<i>M</i>	<i>SD</i>	<i>Min</i>	<i>Max</i>	<i>α</i>
Can Do ^a	1,135	4.80	1.22	1.00	7.00	.89
Effort & Personal Discipline ^a	1,135	5.15	1.36	1.00	7.00	.81
Physical Fitness & Bearing	1,132	5.19	1.54	1.00	7.00	--
Self-Management ^a	1,135	5.26	1.08	1.00	7.00	.75
Working with Others ^a	1,135	5.22	1.19	1.00	7.00	.62
Overall Leadership Potential	1,124	4.72	1.64	1.00	7.00	--
Army-Wide ^a	1,136	5.08	1.09	1.00	7.00	.94

Note. Ratings on PRS range from 1 and 7. PRS ratings from supervisors with a familiarity rating of 1 (“I have had little opportunity to observe this Soldier”) were excluded from analyses. α = coefficient alpha.

^a Ratings composite comprises two or more Army-wide PRS.

Table C.3. Correlations Among PRS Scores in the Full In-Unit Sample

Army-Wide PRS	1	2	3	4	5	6	7
1. Can Do ^a							
2. Effort & Personal Disc ^a	.75						
3. Physical Fitness & Bearing	.58	.63					
4. Self-Management ^a	.78	.75	.60				
5. Working with Others ^a	.77	.77	.60	.75			
6. Overall Leadership Potential	.68	.70	.60	.66	.65		
7. Army-Wide ^a	.93	.89	.72	.90	.89	.75	

Note. Army-wide PRS, $n = 1,120 - 1,136$. . Ratings on PRS range from 1 and 7. PRS ratings from supervisors with a familiarity rating of 1 (“I have had little opportunity to observe this Soldier”) were excluded from analyses. All correlations are statistically significant ($p < .05$, one-tailed).

^aRatings composite comprises two or more Army-wide PRS.

Table C.4. Descriptive Statistics and Internal Consistency Reliability Estimates for the Army Life Questionnaire (ALQ) in the Full In-Unit Sample

Domain / Scale	<i>n</i>	<i>M</i>	<i>SD</i>	<i>Min</i>	<i>Max</i>	<i>α</i>
<i>Retention</i>						
Affective Commitment	1,846	3.65	.78	1.00	5.00	.88
Army Career Intentions	1,846	2.69	1.16	1.00	5.00	.92
Army Fit	1,846	3.94	.68	1.17	5.00	.80
Attrition Cognitions	1,846	1.67	.71	1.00	5.00	.77
MOS Fit	1,846	3.26	.92	1.00	5.00	.93
MOS Satisfaction	1,846	3.57	.88	1.00	5.00	.92
Reenlistment Intentions	1,846	3.02	1.11	1.00	5.00	.80
<i>Achievement/Performance</i>						
Accelerated Development	1,830	.27	.58	0.00	3.00	--
Awards Earned (Weighted) ^a	1,845	4.20	9.44	0.00	55.00	--
Disciplinary Incidents	1,846	.32	.82	0.00	7.00	--
Last APFT Score	1,791	244.06	32.51	105.00	300.00	--

Note. α = coefficient alpha.

^aAwards earned are weighted by the number of promotion points associated with each award according to current Army Enlisted promotion policy.

Table C.5. Correlations Among the In-Unit ALQ Scales in the Full In-Unit Sample

Domain/Scale	1	2	3	4	5	6	7	8	9	10
<i>Retention</i>										
1. Affective Commitment										
2. Army Career Intentions	.59									
3. Army Fit	.76	.57								
4. Attrition Cognitions	-.58	-.48	-.67							
5. MOS Fit	.37	.23	.41	-.31						
6. MOS Satisfaction	.49	.33	.51	-.38	.55					
7. Reenlistment Intentions	.53	.82	.55	-.45	.21	.26				
<i>Achievement/Performance</i>										
8. Accelerated Development	.00	-.01	.01	.00	.00	.02	-.01			
9. Awards Earned (Weighted) ^a	-.03	-.04	.01	-.02	.03	.02	-.04	.17		
10. Disciplinary Incidents	-.13	-.09	-.18	.22	-.07	-.08	-.05	-.04	.01	
11. Last APFT Score	.05	.04	.09	-.12	.02	.02	.02	.13	.07	-.04

Note. $n = 1,775$ - 1,846. Correlations in bold are statistically significant ($p < .05$, two-tailed).

^aAwards earned are weighted by the number of promotion points associated with each award according to current Army Enlisted promotion policy.

Table C.6. Correlations Among the In-Unit PRS in the Validation Sample

Army-Wide PRS	1	2	3	4	5	6
1. Can Do ^a						
2. Effort & Personal Disc ^a	.76					
3. Physical Fitness & Bearing	.58	.65				
4. Self-Management ^a	.81	.80	.64			
5. Working with Others ^a	.78	.81	.63	.78		
6. Overall Leadership Potential	.72	.71	.67	.67	.68	
7. Army-Wide ^a	.93	.90	.73	.92	.90	.77

Note. PRS, $n = 214\text{--}218$. Ratings on PRS range from 1 and 7. PRS ratings from supervisors with a familiarity rating of 1 (“I have had little opportunity to observe this Soldier”) were excluded from analyses. All correlations are statistically significant ($p < .05$, one-tailed).

^aRatings composite comprises two or more Army-wide PRS.

Table C.7. Correlations Among the In-Unit ALQ Scales in the Validation Sample

Domain/Scale	1	2	3	4	5	6	7	8	9	10
<i>Retention</i>										
1. Affective Commitment										
2. Army Career Intentions	.59									
3. Army Fit	.77	.57								
4. Attrition Cognitions	-.52	-.45	-.61							
5. MOS Fit	.42	.27	.43	-.28						
6. MOS Satisfaction	.49	.26	.51	-.31	.52					
7. Reenlistment Intentions	.51	.82	.51	-.41	.26	.23				
<i>Achievement/Performance</i>										
8. Accelerated Development	.00	-.05	.03	-.01	-.02	.08	-.08			
9. Awards Earned (Weighted) ^a	.02	-.06	.05	-.03	.07	.07	-.01	.12		
10. Disciplinary Incidents	-.08	-.05	-.08	.23	-.14	-.06	.00	-.03	-.02	
11. Last APFT Score	.06	.01	.05	-.06	.05	-.02	-.03	.13	.07	-.06

Note. $n = 363\text{--}378$. Correlations in bold are statistically significant ($p < .05$, two-tailed).

^aAwards earned are weighted by the number of promotion points associated with each award according to current Army Enlisted promotion policy.